TRANSPORTATION RESEARCH RECORD 1091

Traffic Flow Theory, Characteristics, and Highway Capacity

TRB
TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1986

Transportation Research Record 1091
Price $16.00
Editor: Catherine Nizharadze
Compositor: Joan G. Zubal
Layout: Theresa L. Johnson

mode
1 highway transportation

subject area
55 traffic flow, capacity, and measurements

Transportation Research Board publications are available by ordering directly from TRB. They may also be obtained on a regular basis through organizational or individual affiliation with TRB; affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Traffic flow theory, characteristics, and highway capacity

(Transportation research record, ISSN 0361-1981 ; 1091)
Reports prepared for the 65th annual meeting of the Transportation Research Board.
1. Traffic flow—Congresses. 2. Highway capacity—Congresses. I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 1091 380.5 s 87-1536
[HE336.T7] [388.3'142]
ISBN 0-309-04111-2

Transportation Research Record 1091

The Transportation Research Record series consists of collections of papers on a given subject. Most of the papers in a Transportation Research Record were originally prepared for presentation at a TRB Annual Meeting. All papers (both Annual Meeting papers and those submitted solely for publication) have been reviewed and accepted for publication by TRB's peer review process according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The views expressed in these papers are those of the authors and do not necessarily reflect those of the sponsoring committee, the Transportation Research Board, the National Research Council, or the sponsors of TRB activities.

Transportation Research Records are issued irregularly; approximately 50 are released each year. Each is classified according to the modes and subject areas dealt with in the individual papers it contains. TRB publications are available on direct order from TRB, or they may be obtained on a regular basis through organizational or individual affiliation with TRB. Affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Contents

Further Analysis of the Flow-Concentration Relationship

FRED L. HALL AND MARGOT A. GUNTER

On the basis of inspection of time-traced plots of daily flow and concentration data for a freeway, additional support is provided for some tentative new ways of looking at the relationship between these variables. Occupancy data are used to directly measure concentration, rather than converting to density. The overlaid, daily time traces help to make clear the nature of operations and of transitions between congested and uncongested regimes. Four principal conclusions are supported by these plots. First, the underlying inverted V-shape for the flow-occupancy relationship found earlier at a single lane and station on the same freeway has been confirmed at several locations for both of the nonshoulder lanes. Second, for the shoulder lane, it is not possible to determine from these data whether the inverted V-shape correctly describes the relationship. Third, definite differences exist in the parameters of the flow-occupancy relationship that appear to be attributable to lane and location. Fourth, there is additional support for the finding reported in an earlier paper of the authors (*Transportation Research*, Vol. 20A, 1986) that discontinuous relationships are not necessary to describe the data obtained from freeway operations. Better sense may be made by assuming continuous relationships and trying to explain sparseness of data by the nature of operations on the facility.

Despite some 50 years of research on the relationships between speed, flow, and vehicular concentration, disagreement still exists about what exactly occurs on freeways. An extensive test (using nearly 50 different data sets) of single- and two-regime models, as derived from first principles, has indicated that the match between theory and data is not very good *(1)*. Therefore, a different approach is taken in this paper: the data are inspected closely to determine the form of relationship that is suggested.

In a previous paper, the authors examined flow-occupancy data for one lane at one location on an expressway *(2)*; on the basis of that examination, three conclusions were reached about the relationship. The purpose of this paper is to determine if those same conclusions are applicable to other lanes and locations. That analysis was directed at the conceptual problem caused by gaps in observed data patterns, and suggested that it is not necessary to construct discontinuous functions to account for those gaps. The three main conclusions of that paper are summarized briefly here because they provided the starting point for the current analysis.

First, there are advantages to looking closely at daily data to discern operating relationships, rather than relying uncritically on scatter diagrams of all available data. In particular, it was found helpful to utilize time-connected traces of the daily record of operations. Second, inspection of daily time-traced plots showed a variety of types of transitions from uncongested to congested flow occurring, but the combined result of these types of transitions led to an appearance of sparsity, or even gaps in the data. It was therefore hypothesized that (2, p.210)

> The nature of the data that are collected at any particular freeway location depends as much on the specifics of the location as on underlying relationships. In particular, there will be an absence of data for particular parts of the relationship if a queue backs into the location while flow is lower than capacity.

Third, because of this explanation for areas of sparse data, arguments for a discontinuous flow-occupancy (or flow-density) curve do not appear to be convincing. An inverted V-shape for a continuous curve appears to be the most representative shape, given the data examined.

The problem with the latter two conclusions is that they are based on data from only one lane (the median, or left-most, lane), at one station on the roadway (4 km upstream from a bottleneck). In this paper, flow-occupancy data for other lanes and locations along the roadway are examined to determine the extent to which those two conclusions are affected by location along the roadway—particularly with reference to entrance ramps—and by lane.

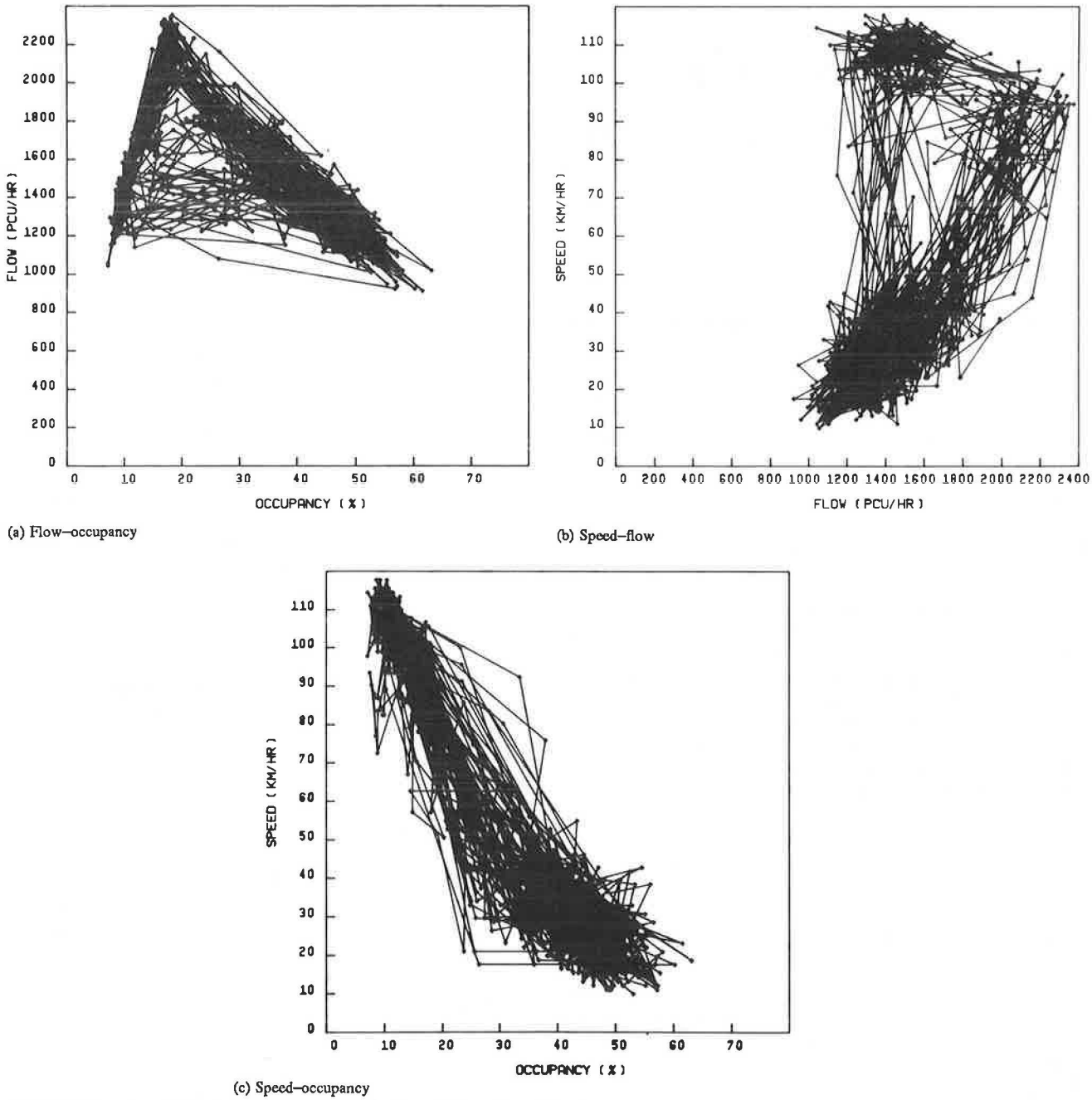
In particular, three questions are addressed in this paper:

1. Is the inverted V-shape observed at other stations and for other lanes?
2. If it is, does it appear to have similar or different parameters at different locations on the road?
3. Are the patterns seen at the several lanes and stations consistent with the idea of a continuous curve, or does a discontinuous curve appear to be appropriate?

It should be recognized from the start that any conclusions must still be tentative because all of the data come from the same freeway control system. It will remain to be seen if the relationship is different elsewhere.

The flow-occupancy relationship was selected for consideration over the speed-flow or speed-occupancy relationships because, in the authors' initial analysis of one lane and station, it provided the clearest distinction between congested and uncongested regimes (Figure 1). The flow-occupancy relationship therefore offered the most promise for being able to clearly specify the nature of the relationship, and subsequently identify the nature of differences in it between stations and lanes on a roadway. Occupancy is used rather than density for two reasons. First, it is the variable directly measured by the freeway management system and is a point or very short section measurement, which corresponds well with the

F. L. Hall, Department of Civil Engineering, McMaster University, Hamilton, Ontario L8S, 4L7, Canada. M. A. Gunter, IBI Group, 240 Richmond St. West, Toronto, Ontario M5V 1W1, Canada; formerly with McMaster University.



(a) Flow–occupancy

(b) Speed–flow

(c) Speed–occupancy

FIGURE 1 Overlaid time-traced plots for the Station 4 median lane.

other data collected by the system. Second, any conversion of this measurement to density (the number of vehicles per kilometer) introduces a large amount of scatter, as can be observed from Koshi et al. (*3*, Figure 1).

The idea that the relationships for different lanes and locations are not the same is not new. Others who have demonstrated or mentioned these differences include the 1965 *Highway Capacity Manual (4)*, Ceder and May *(5)*, Mahabir *(6)*, and Hurdle and Datta *(7)*. However, it is logical to expect that more precise specification of the traffic flow relationships for different locations will be necessary as freeway management systems become increasingly important and widespread.

The first section of this paper contains a description of the data set used in the study. The second section provides an account of the computerized procedures used to reduce the set to a form suitable for analysis. The results of the analysis are presented in the third section. Discussion of the implications and possible interpretation of the results are given in the next section. The final section contains the conclusions and suggestions for future work.

DATA

The data used in this study come from the freeway control system on Ontario's Queen Elizabeth Way between the cities of Oakville and Toronto. A 5-km section of this roadway experiences congestion in the eastbound direction each weekday

morning and has therefore been equipped with a freeway management system that collects traffic flow data at nine locations or stations [described in more detail by Case and Williams *(8)*]. At each lane at each station (lane-station), the following were compiled by the freeway management system for 5-min time intervals: occupancies at the upstream and downstream loops, number of vehicles longer than 7.6 m [a passenger car equivalency of 2 was used for these long vehicles to convert the total volume to passenger car units (pcu's)], total number of vehicles, and average speed. In addition, a log was kept by the system operator, noting poor weather or incidents such as accidents or stalled vehicles.

For this analysis, 8 months of data were available, collected in 1978 and 1979. The 5-km section of the freeway system is shown in Figure 2, along with the locations of some of the data collection stations. The primary bottleneck is just downstream of the entrance ramp merge at Highway 10. (Unfortunately, data were not collected there.) This is a secondary constriction at the entrance ramp from Mississauga Road, which also has heavy entrance-ramp volumes.

The choice of stations for analysis was important because the effect of location is a major component of the study. The primary criterion was that the traffic conditions be as different as possible. The second criterion was that geometric conditions should be as close to ideal as possible. These two criteria led to the choice of three stations: Stations 9, 7, and 4.

Station 9 is located immediately upstream of the bottleneck and the Highway 10 entrance ramp and is on a slight vertical curve. It experiences congestion for the longest duration of all locations along the section.

Station 7 is situated about 1.6 km upstream of Station 9, with one exit ramp intervening between them. It has almost ideal geometric conditions, but is immediately downstream of a bridge. Just before the bridge is the end of the merging lane for the entrance ramp from Mississauga Road.

Station 4 has good geometrics. It is about 3.9 km from Station 9, and there are two entrance ramps and one exit ramp between it and Station 7.

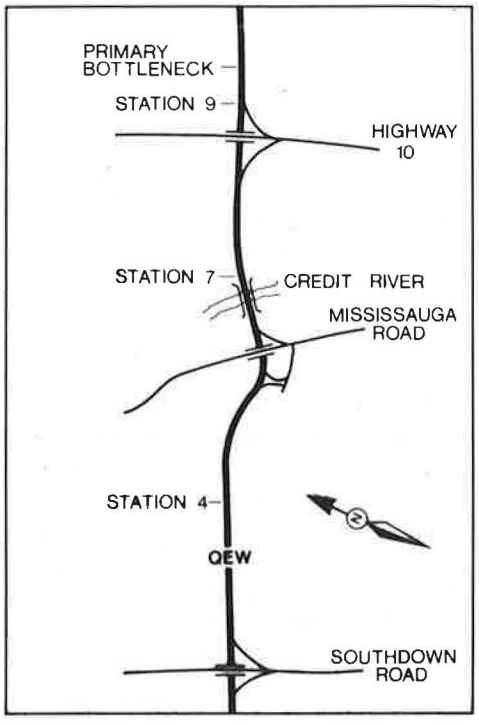


FIGURE 2 Schematic map of freeway system (not to scale, only eastbound ramps shown).

DATA REDUCTION

After the locations and lanes were chosen, the raw data had to be reduced to a set that was appropriate for analysis, and displayed in a form that would be useful for subsequent interpretation. Specifically, the goal for the data-reduction stage was to produce for each lane-station a figure containing overlaid time-traced plots of the nontransitional data for those days that represented ideal conditions.

Because this analysis was intended to compare ideal relationships, days were rejected if weather had been poor or if incidents (stalled vehicles, accidents) had occurred. This reduced the data to 72 days of operation.

The next step was to distinguish between the two regimes of operations for the data for each lane. A critical occupancy value, at which flow was at a maximum, was identified. The choice of this value was somewhat subjective, but could be made with reasonable confidence based on examination of the overlaid daily plots and of plots of average values, following procedures described by Hall, Allen, and Gunter *(2)*.

This procedure worked well with all the middle and median lane plots, but was not suitable for the shoulder lane. In the shoulder lane a peak flow could not easily be distinguished, and the averaging procedures did not help. Therefore, a subjective, and admittedly somewhat arbitrary, value was selected for the critical occupancy to indicate the value at which congested operation began based on the appearance of the plot.

After a critical occupancy had been established, it was possible to identify and reject inappropriate points. For the authors' previous paper *(2)* a manual technique was used, which involved examining each day's time-traced data separately. To analyze all nine lane-stations efficiently, a computerized procedure was developed that used the same criteria.

These criteria defined four categories of rejection for individual data points:

1. Equipment malfunction,
2. Serious doubt as to the data's numerical validity,
3. Inconsistency with the equilibrium relationship, and
4. Operation in transition between the two branches of the curve.

Each of these categories of rejection will be discussed further.

Most equipment malfunctions had been identified automatically by the system. Two other types were found and rejected: zero recorded for flow rate, speed, or occupancy; and more than a 40 percent difference between upstream and downstream occupancy (which usually differed by less than 20 percent in these data).

The second criterion dealt with serious doubt about particular observations. This was characterized by a lack of reliable

data in a day, defined as less than 0.5 hour as data between equipment failures, or less than 0.5 hour of data in total.

The third criterion that led to rejection of data was an occupancy or speed value that was obviously inconsistent with the equilibrium relationship. Speed was also used because there were a few cases in which it was the only factor out of line, but such a point should also be questioned. Inconsistency was defined on the basis of values greater than two standard deviations away from the average for their flow range, or consistently between one and two standard deviations removed in the direction of transition to the other regime.

The fourth type of rejection, points in transition between regimes, deserves more explanation. The basic rationale is that it is illogical to include data that are in transition between regimes when one wants to compare only the stable operation in each branch of the curve. [The importance of this is also discussed by Payne *(9)*.] Two criteria were established, both of which had to be met to define a point as being in transition. The first was that a data point must have speed or occupancy greater than one standard deviation from the average value in the direction of the other regime, and must continue to move away from its present regime in the next time interval. The second criterion was that the flow had to be less than a critical value (generally 2,000 pcu/hr for the median lane) to be rejected because it was judged to be in transition. This second criterion was added for the middle and median lanes because of the uncertainty about the location of the function at high flow rates.

The final step in the data reduction for all lane-stations was to check the computerized result manually. This was accomplished by plotting the overlaid time-traced nontransitional data and visually checking the plot for any obvious outliers, transitional points, or relationships with inconsistent structures. The first two anomalies were corrected by manually searching the data and deleting them the dozen or so times they occurred. Fortunately, the third irregularity did not occur.

ANALYSIS

The data-reduction procedure thus provided nine plots: one for each lane at each of the three locations. (See Figure 3; note the following in this figure: top row, Station 9; middle row, Station 7; bottom row, Station 4; left column, median lane; center column, middle lane; right column, shoulder lane.) These plots are intended to represent nontransitional (or equilibrium) operation. The analysis consists primarily of visual inspection and comparison of these plots in order to answer the three questions raised at the beginning of this paper:

1. Is the same general shape (an inverted V-shape) observed at all lanes and stations?
2. If the shape is the same, are the parameters (particularly maximum flow rates and critical occupancies) similar?
3. Are areas of sparse data (or discontinuities) present at all locations, and, if so, are they related in a sensible way?

The answer to the first question is that there is an overall consistency in the general shape of the relationships for two lanes, but not for the shoulder lane. The six plots for the middle and median lanes all show relationships similar to that at Station 4, median lane (Figure 1a): a well-defined uncongested regime, a sharp peak in flow, and a congested regime with a large amount of scatter in the data. In the three shoulder lane plots, the maximum observed flow rates are attained in the congested regime, and no distinct peaks are observed.

Qualitative comparisons were made of the parameters for these relationships. Six comparisons were made: one for each of the median, middle, and shoulder lanes comparing the relationships at the three stations along the roadway to find the effect of location; and one for each of the three stations, comparing across the three lanes. A summary figure was drawn for each combination to be compared, which indicates the outline of the general location of the data points for the relevant lane stations.

For the median lane, distinct differences exist between the flow-occupancy relationship at Station 7 on the one hand and Stations 4 and 9 on the other (Figure 4). One possible explanation for this behavior is that Station 7 is located in a secondary bottleneck (secondary in the sense that there is another one farther downstream that causes a queue as far back as Station 7 or farther.) Hurdle and Datta *(7)* have suggested that speed-flow relationships may differ in a bottleneck; the flow-occupancy curve may also be affected in some way. The maximum flows observed at Station 7 are 200 pcu/hr higher than those at the other locations. The occupancies corresponding to these maximum observed flows are also 2 to 4 percent higher, which necessarily follows from the higher flows and the similarity of the nearly linear relationship in the uncongested regime over the three locations. This similarity does not occur in the congested regime, where two separate curves operate, one for Stations 4 and 9 and one for Station 7. Both curves have a similar shape, but it would appear that after the road is congested, the mean flow rates for the secondary bottleneck are consistently 400 pcu/hr higher at any given occupancy.

The middle lane data exhibit a different pattern in that the three stations appear to have consistent speed-occupancy patterns (Figure 5). For both the uncongested and congested regimes, the data for all three stations lie almost directly on top of each other, clearly suggesting that a single relationship can represent them all well. No obvious reasons for this difference from the median lane occur to the authors.

Three observations can be made from consideration of the plot for the shoulder lanes (Figure 6). First, the uncongested regimes for Stations 4 and 7 coincide, but Station 9 has generally lower flows for any given occupancy. (The Station 9 shoulder lane is heavily affected by an entrance ramp merging immediately downstream.) Second, in the congested regime the data for the three stations are again separated, somewhat as in the median-lane comparison. For any given occupancy, Station 7 has the highest flows, Station 9 the lowest, and Station 4 is in between. Third, all shoulder lane plots exhibit higher maximum flows in the congested regime than in the uncongested regime, which is contrary to their definitions. (This may be a consequence of decreased flows on the metered ramps, leading to increased main-line flows as the system becomes more congested.)

Two plausible interpretations of the flow-occupancy relationship are consistent with the second and third of these observations. Either there is an underlying relationship in the shoulder lane similar to that in the middle and median lanes,

with a distinct peak in the relationship even though it has not been observed in the data, or the relationship in this lane is fundamentally different. In the first case, the gap in the data, or the unobserved portion of the curve, includes all of the operation around capacity. The alternate explanation implies that what the authors have termed transitions in fact represent operations at the top of those curves, near the capacity for shoulder lanes.

The resolution of this problem may be aided by a comparison across lanes at each station (Figure 7). The first observation is that in the uncongested regime the slope of the relationship becomes increasingly steep from the shoulder to

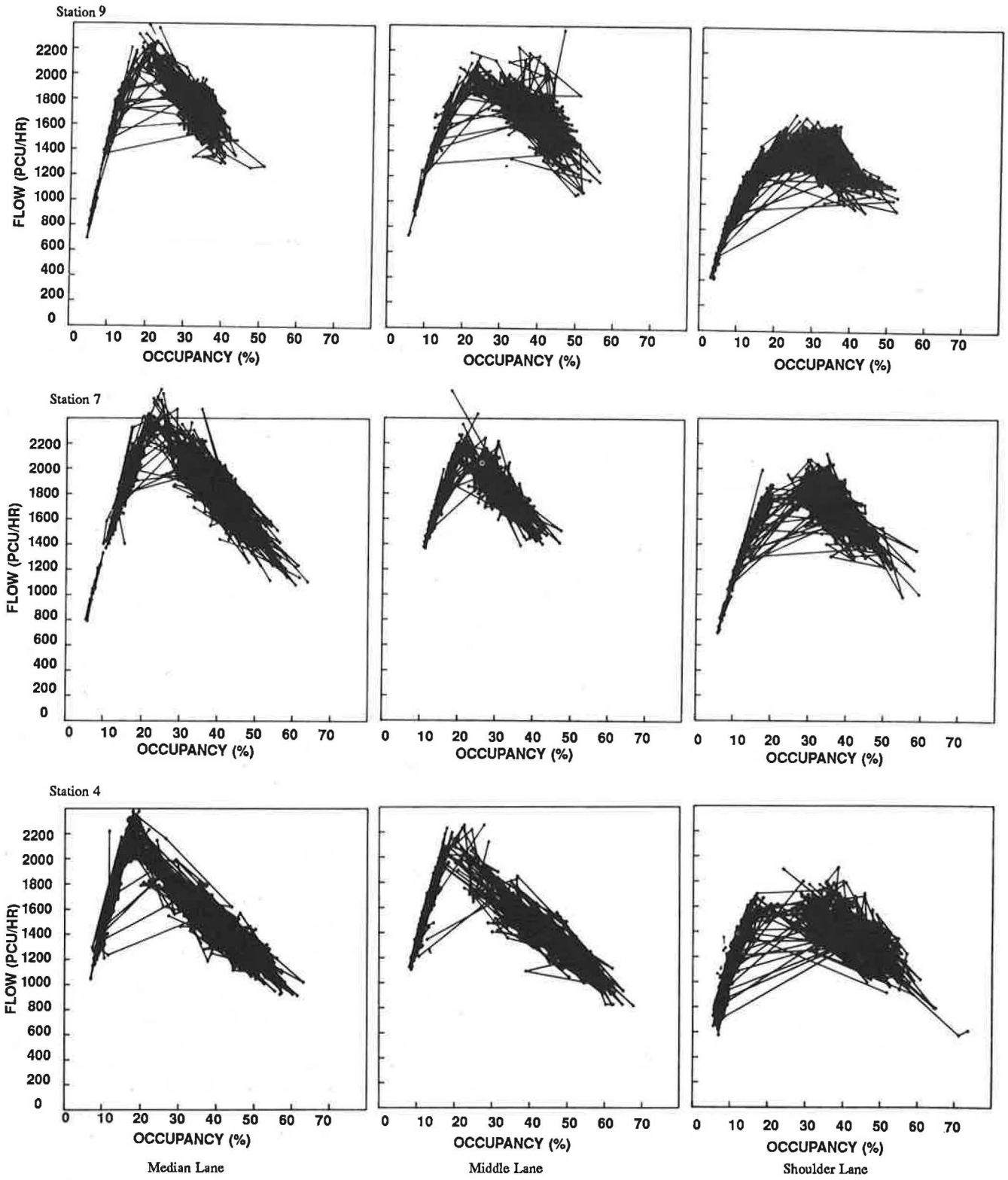


FIGURE 3 Nontransitional, overlaid time-traced plots for three lanes and three stations.

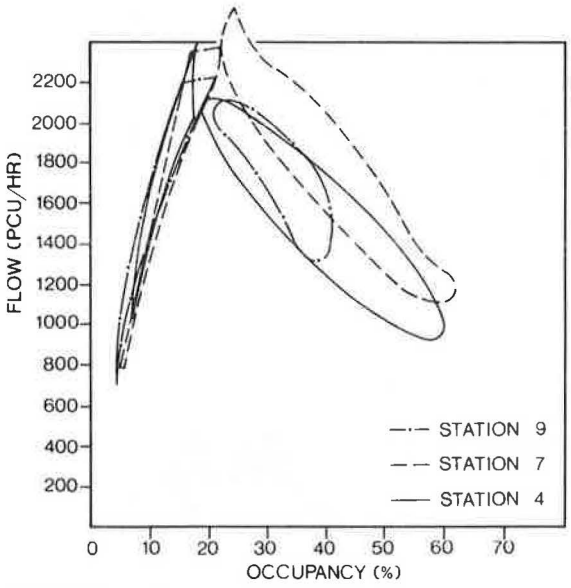


FIGURE 4 Comparison of median lanes across stations.

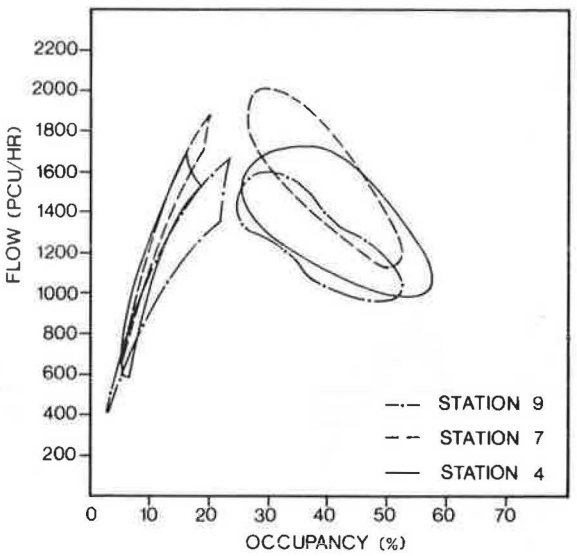


FIGURE 6 Comparison of shoulder lanes across stations.

the median lane. This suggests that for a given occupancy, drivers in the median lane will manage a higher flow rate; or conversely, for a given flow rate, median-lane drivers will accept higher concentrations of vehicles. This characteristic of driver behavior comes as no surprise.

The second observation pertains to the congested regime. At Station 4, the data for all three lanes coincide very closely. At Station 7, the data for the shoulder and middle lanes coincide, but only about one-half of the data for the median lane falls in the same area, with the remainder shifted up and to the right. At Station 9, the median and middle lanes follow a consistent pattern, but the data for the shoulder lane are clearly shifted down and to the left.

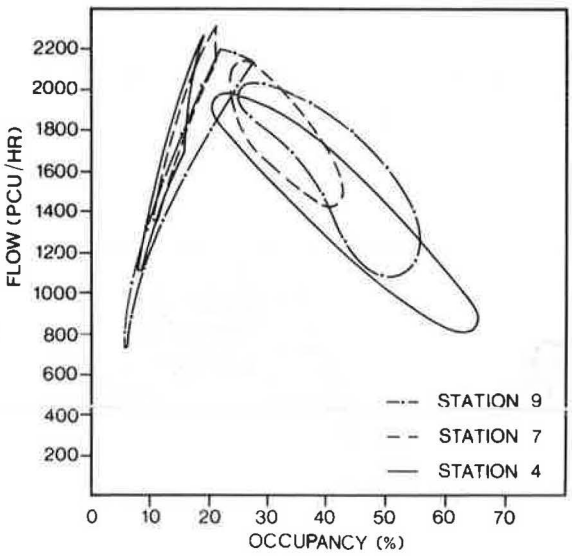


FIGURE 5 Comparison of middle lanes across stations.

The net result is to leave the question about the shape of the shoulder-lane plot unresolved. The inverted V-shape found in nonshoulder lanes may not apply to the shoulder lane, even ignoring Station 9 as being too close to an entrance ramp to include in any generalization. Whether it is the same curve depends ultimately on the missing data, or the gaps in the data, which can only be resolved by looking at data from other freeways and locations.

The third question addressed by the analysis dealt with the question of discontinuities in the functions. The authors have suggested that the discontinuities others have hypothesized for relationships describing freeway data are unnecessary for understanding the nature of operations. Further, the wide variety of different functional forms that have been calculated by proponents of discontinuous two-regime models [e.g., Payne *(9)*, Ceder *(10)*, and Easa *(11)*] may well have obscured the systematic variation that is to be expected from the queueing systematic variation that is to be expected from the queueing process that takes place on freeways.

One should expect to find different apparent discontinuities in the data, depending on the location relative to high-volume entrance ramps and on the flow rates on the main line when a queue reached that location. The results support this argument. For example, it is clear that the concentration of data in the congested regime for Station 4 occurs at flows lower than for the other stations (Figure 3). This is reasonable because the metered entrance ramps at Mississauga Road add 1,000 or so vehicles per hour to the flow. Station 4, when congested, must move fewer vehicles than does Station 7 or 9. The consequence of this for the data is that Station 4 will have sparse data in the upper portion of the congested regime in a flow-occupancy curve. In a speed-concentration curve such as other researchers have focused on, the sparse data will occur somewhere in the middle of the curve, which may lead one to infer the need for a discontinuous function. However, that sparsity of data need not

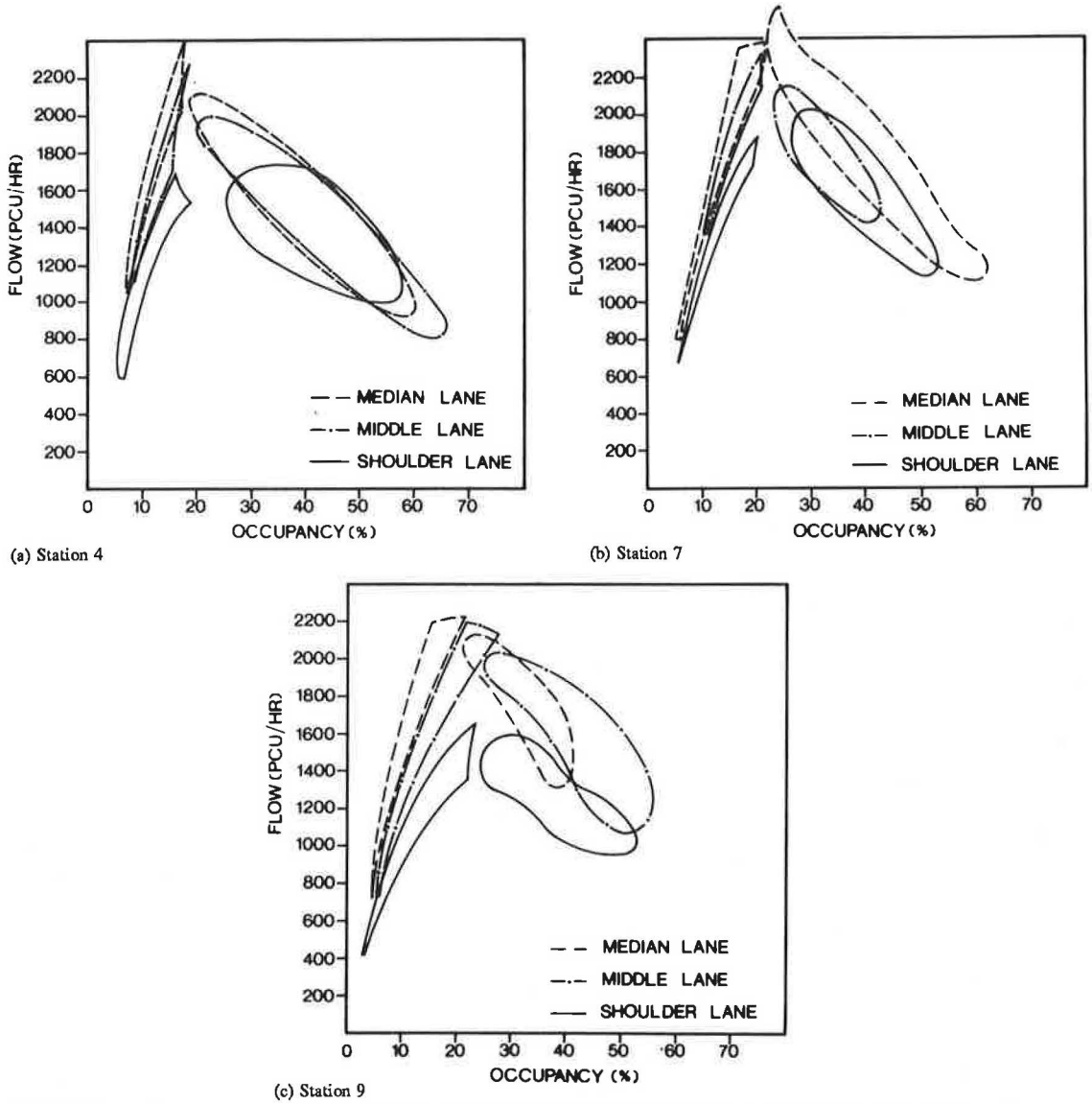


FIGURE 7 Comparison across lanes at each of three stations: Station 4, Station 7, and Station 9.

imply a discontinuity of the function, nor a different functional form.

DISCUSSION OF RESULTS

Two important results arise from the analyses, in addition to the support provided for the authors' earlier arguments against the necessity for discontinuous functions. First, the inverted V-shape for the flow-occupancy relationship is found for both of the nonshoulder lanes at all stations examined. Whether the shoulder lane conforms to this same pattern is still an open question. Second, the parameters of the flow-occupancy relationship for a freeway differ across stations along the roadway. In particular, maximum flow rates were higher at Stations 7 than at 4 and 9 (median and shoulder lanes), and this appears to be reflected in a slightly shifted curve in the congested regime. The analysis was begun with the hope of demonstrating that the results obtained earlier for one lane at one station could be generalized. The current results suggest that simple generalizations will be inadequate; therefore, in this section the simplest way the authors can think of to deal with the differences observed is considered.

One relatively simple explanation was rejected, namely, that the differences arise from differences in the geometric characteristics of the highway. Lane width across all four stations is almost identical. The shoulders are wider at Station 4 than at Station 7 or 9, but that would suggest that Station 4 should have the highest maximum flows. Likewise, Station 4 is on a straight level section of highway, whereas Station 7 is on a small grade and at the end of a horizontal curve. Again, this would suggest that Station 4 should have the higher flows. Station 9 is on a vertical curve, so it is surprising that in several respects it is similar to Station 4. Geometric characteristics do not appear to provide the explanation for the differences across stations.

With the obvious solution discounted, what is left is the idea that the high flow rates and shifted congested regime at Station 7 are due to the station's location in a secondary bottleneck. This idea about bottleneck flows deserves a careful discussion. The authors are not convinced that it is the explanation for the observed differences, but any other possibilities have been ruled out.

The origins of the hypothesis come from a paper by Hurdle and Datta, in which they observed some surprising results in speed-flow data *(7)*. They hypothesized that very high flows (above 1,850 pcu/hr/lane) are associated with a slight drop in speed, and that these speeds and flows occur in a bottleneck when the vehicles are "being discharged from an upstream queue."

However, in the Queen Elizabeth Way system careful examination of daily data from the secondary bottleneck and from Station 6 upstream reveals that Station 7 becomes congested because of a queue from the primary bottleneck before a queue can form from the secondary bottleneck. It is therefore not possible to be certain that the traffic flow relationships are different in the congested regime because they are in a secondary bottleneck location that has been fed by a queue. Nevertheless, at Station 7 there is a situation in which it is suspected, from the geometry of the situation, that a secondary bottleneck exists. The extra queue upstream of Station 7 is caused by heavy mainline flow merging with two heavily-used entrance ramps from Mississauga Road (which are metered at rates of approximately 12 vehicles per minute during this period), thereby making Station 7 act in part as a bottleneck.

Two possible explanations are offered for the observed differences in the flow-occupancy relationship in this situation. The first is the presence of metered ramps. The logic for ramp-metering systems is to control entering traffic to maintain a relatively smooth traffic flow downstream. It is entirely possible that these results are an eloquent demonstration of just how much ramp-metering systems have accomplished. The data in this paper, and in Hurdle's, come from a functioning freeway management system. It may well be that operating characteristics have changed, particularly around capacity and in the congested regime, in which ramp-metering systems are most often used. In other words, there is the possibility of a different relationship in the congested regime because of intervention by traffic engineers.

Another possible explanation is that any time the congested flow is fed by an upstream queue, the flows will be higher than if it is not. In other words, the shift in the flow-occupancy function occurs because the vehicles are coming from an upstream queue. This extends Hurdle and Datta's reasoning to a different relationship than they had proposed, but appears to be in agreement with their suggestions. The underlying mechanisms behind this behavior are certainly not obvious. Perhaps, after experiencing stop-and-go conditions combined with merging vehicles, drivers will take advantage of an uninterrupted stretch of roadway. It may therefore be a measure of the drivers' increasing frustration that higher flow rates and occupancies are found at Station 7 than at Station 4 (which has better geometrics, but has not been preceded by a merging section).

Whichever explanation is correct, the generalization of these results is shown in Figure 8. In the uncongested regime, there is a well-defined, nearly linear function, but with lower slopes and lower maximum flows as one moves from the median to the shoulder lane. In the congested regime, there is a broad band of possible data points, which should not be represented by a single line. Despite the scatter, in a bottleneck the curve is shifted up and to the right. The authors are unable to draw the shoulder lane plot near its capacity.

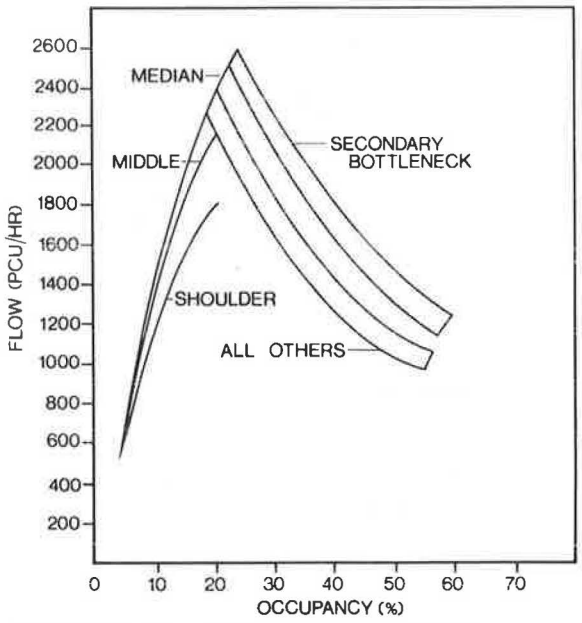


FIGURE 8 Generalization of observed flow-occupancy relationship.

CONCLUSIONS

In this paper, additional support has been provided for some tentative new ways of looking at relationships between variables describing freeway operations. The main conclusions are as follows:

1. An underlying inverted V-shape for the flow-occupancy relationship has been confirmed at various locations along a freeway for both of the nonshoulder lanes.
2. For the shoulder lane, it is not clear whether the inverted V-shape holds or whether there is a different, inverted U-shape relationship, such as that shown in *Transportation Research Circular 281* (*12*, Figure 3-3).
3. Definite differences have been found in the parameters of the flow-occupancy relationship that appear to be due to lane and location. In particular, the following trends were noted.

   a. Locations in a secondary bottleneck exhibit higher maximum uncongested flows and occupancies and higher flows at all occupancies in the congested regime.
   b. Within a single lane, the uncongested regime is well defined, but the flow-occupancy slope becomes increasingly steep moving across lanes toward the median lane.
   c. Proximity to an entrance ramp decreases flows at all occupancies in the shoulder lane.

4. Discontinuous relationships are not necessary to describe the data obtained from freeway operations. Better sense may be made by assuming continuous relationships, and trying to explain sparseness of data by the nature of operations on the facility.

These conclusions are admittedly tentative. To verify or discredit them, data need to be obtained from other freeway systems. In particular, data are needed from various locations upstream from a bottleneck, locations that experience different flow rates at the time they switch to congested flow. In addition, data on the shoulder lane in the bottleneck would help to establish the nature of the flow-occupancy relationship for that lane.

These findings, if confirmed, have important implications in many areas of theoretical and practical traffic engineering. The most important of these is that one needs to pay attention to the location of the data acquisition in order to make sense of the results. As freeway management systems become more common, this requirement becomes increasingly important.

ACKNOWLEDGMENTS

The data used in this analysis were provided by the Ontario Ministry of Transportation and Communications, and were made workable on the McMaster University computer facilities by Geddes Mahabir several years ago. Funding for this project was through the Natural Sciences and Engineering Research Council's Undergraduate Summer Research Award programme. The authors should also acknowledge the helpful and thought-provoking comments of two reviewers of a previous paper.

REFERENCES

1. A. Ceder. *Investigation of Two-Regime Traffic Flow Models at the Micro- and Macro-scopic Levels*. Ph.D. dissertation. University of California, Berkeley, 1975.
2. F. L. Hall, B. L. Allen, and M. A. Gunter. "Empirical Analysis of Freeway Flow–Density Relationships." *Transportation Research*, Vol. 20A, Elmsford, N. Y., 1986, pp. 197–210.
3. M. Koshi, M. Iwaski, and I. Okhura. "Some Findings and an Overview on Vehicular Flow Characteristics." *Proc., 8th International Symposium on Transportation and Traffic Flow Theory, 1981*. University of Toronto Press, Toronto, Ontario, Canada, 1983, pp. 403–426.
4. *Special Report 87: Highway Capacity Manual—1965*, HRB, National Research Council, Washington, D.C., 1966, 411 pp.
5. A. Ceder and A. D. May. "Further Evaluation of Single- and Two-Regime Traffic Flow Models." In *Transportation Research Record 567*, TRB, National Research Council, Washington, D.C., 1976, pp. 1–15.
6. G. P. Mahabir. *Speed, Flow, and Capacity Relations on Multilane Highways*. M. Eng. thesis. Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada, 1981.
7. V. F. Hurdle and P. K. Datta. "Speeds and Flows on an Urban Freeway: Some Measures and a Hypothesis." In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp. 127–137.
8. E. R. Case and K. M. Williams. "Queen Elizabeth Way Freeway Surveillance and Control System Demonstration Project." In *Transportation Research Record 682*, TRB, National Research Council, Washington, D.C., 1979, pp. 84–93.
9. H. J. Payne. "Discontinuity in Equilibrium Freeway Traffic Flow." In *Transportation Research Record 971*, TRB, National Research Council, Washington, D.C., 1984, pp. 140–146.
10. A. Ceder. "A Deterministic Traffic Flow Model for the Two-Regime Approach." In *Transportation Research Record 567*, TRB, National Research Council, Washington, D.C., 1976, pp. 16–32.
11. S. M. Easa. "Selecting Two-Regime Traffic-Flow Models." In *Transportation Research Record 869*, TRB, National Research Council, Washington, D.C., 1983, pp. 25–36.
12. R. P. Roess, et al. "Freeway Capacity Procedures." In *Transportation Research Circular 212: Interim Materials on Highway Capacity*, TRB, National Research Council, Washington, D.C., 1980, pp. 151–266.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Passenger Car Equivalents for Trucks on Level Freeway Segments

RAYMOND A. KRAMMES AND KENNETH W. CROWLEY

The term passenger car equivalent (PCE) was introduced in the 1965 *Highway Capacity Manual*. Since 1965, considerable research effort has been directed toward the estimation of PCE values for various roadway types. However, at present, there is neither a commonly accepted nor clearly defined theoretical basis for the concept of passenger car equivalency. Two components of a theoretical basis for equivalency are defined in this paper: *(a)* that the basis for equivalence should be the parameters used to define level of service for the roadway type in question, and *(b)* that the PCE formulation should be expressed in terms of variables that reflect the relative importance of three factors that contribute to the overall effect of trucks on that roadway type. The three factors are *(a)* trucks are larger than passenger cars, *(b)* trucks have operating capabilities that are inferior to those of passenger cars, and *(c)* trucks have a physical impact on nearby vehicles and a psychological impact on the drivers of those vehicles. The two components of the theoretical basis were used to evaluate the merits of three approaches to estimating PCEs for level freeway segments: *(a)* the constant volume-to-capacity ratio approach, *(b)* the equal-density approach, and *(c)* the spatial headway approach. It was concluded that the spatial headway approach was appropriate for level, basic freeway segments, and a PCE formulation expressed in terms of headway measurements was derived.

Examined is the estimation of passenger car equivalents (PCEs) for trucks on level freeway segments. PCEs are used in capacity analysis procedures to convert mixed traffic stream volumes into equivalent passenger-car-only volumes. Level freeway segments are important because they are prevalent in urban areas, where traffic congestion is most common.

The need for additional consideration of this topic stems from two problems. First, the research effort to estimate PCEs for trucks on level freeway segments has been limited. Second, there is neither a commonly accepted definition of equivalence nor a clearly defined theoretical basis on which to derive PCE formulations.

The term PCE was first used in the 1965 *Highway Capacity Manual* (HCM) *(1)*, and since its publication at least 12 studies have documented approaches to estimating PCEs. Most of the research applied to two-lane or multilane highways *(2–9)*. Considerable effort has also been expended to update PCE values for specific grades on freeway facilities *(10)*. However, of the three studies applicable to level freeway segments, two were limited to specific sites: the Baltimore Harbor Tunnel *(11)* and the M4 motorway in London, England *(12)*. Only a recent study by the Institute for Research (IFR) involved a broad-based data collection effort at 11 level freeway sites in 4 urban areas in the United States *(13)*. However, IFR estimated PCE values for use in a highway cost allocation study and not specifically for capacity analysis purposes, and the two uses may not be compatible *(13)*.

Roess and Messer *(14)*, co-editors of the 1985 edition of the HCM, reviewed most of the studies just referenced and concluded the following *(15)*:

> Because of the wide variance in pce philosophies adopted by researchers, it is difficult to directly compare numerical results. Unfortunately, there was no uniform understanding of what a pce *meant* before the above studies were undertaken, and indeed the *intended use* of results also varied.

Roess and Messer *(14)* identified three approaches that "appear to have direct relevance to highway capacity analysis":

1. The constant volume-to-capacity ratio approach,
2. The equal-density approach, and
3. The spatial headway approach.

The PCE value for trucks on level freeway segments in the 1985 *HCM (15)* is based on the study by IFR, which used a spatial headway approach *(13)*. However, Roess and Messer indicate *(14)*:

> Unfortunately, it will not be possible to reconcile these three approaches as new capacity techniques are developed in anticipation of a third edition of the *Highway Capacity Manual*. The data bases are incompatible, and do not allow revision of the results of these studies into a single format. Thus, elements of all three principles will survive into new techniques.

Evaluated in this paper are the merits of these approaches for level freeway segments. First, two principles are defined as components of a theoretical basis for the concept of passenger car equivalency. Then these principles are used as the basis for the evaluation of the three approaches and for the derivation of the PCE formulation used by IFR *(13)*. Finally, a more sophisticated headway-based formulation, which may be more appropriate for highway capacity analysis, is identified.

THEORETICAL BASIS FOR PASSENGER CAR EQUIVALENCY

Two basic principles should be applied to the estimation of PCE values for any of the roadway types identified in capacity analysis procedures. The first principle links the concept of passenger car equivalency to the level of service (LOS) concept. The second principle emphasizes the consideration of all

R. A. Krammes, Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843. K. W. Crowley, Institute for Research, 257 South Pugh St., State College, Pa. 16801.

factors that contribute to the overall effect of trucks on traffic stream performance.

Role of PCEs in Capacity Analysis

Highway capacity analysis procedures are based on the LOS concept, which correlates the driver's perception of operating conditions with traffic flow parameters such as speed or density. According to Roess, "The Level of Service Concept is defined to be quality of service as defined by the highway user" *(16)*. The 1965 *HCM* described LOS as "a qualitative measure of the effect of a number of factors, which include speed and travel time, traffic interruptions, freedom to maneuver, safety, driving comfort and convenience, and operating cost" *(1)*.

Operating conditions on a highway are divided into six levels, A through F. Each level represents a limited range of operating conditions and is defined in terms of minimum or maximum values of traffic flow parameters that reflect the driver's perception of the quality of service provided by the facility. Because, for each roadway type, the combination of factors that influence the driver's perception of conditions is different, the parameters that are used to define LOS also differ *(16)*.

The capacity analysis procedures are calibrated for a specific set of ideal conditions, one of which is that the traffic stream contains only passenger cars. Adjustments are made for deviations from those ideal conditions; the adjustment factor for the presence of trucks is based on PCEs. This adjustment factor correlates the flow rates of passenger cars only and of mixed traffic streams that are equivalent in terms of the driver's perception of the quality of service. Because the parameters that are used to define LOS reflect the factors that influence the driver's perception, the same parameters should be used to compare passenger cars and trucks and to estimate PCEs. Roess and Messer support this contention when they state, "As Level of Service criteria for capacity analysis are based upon performance parameters, it is logical that PCE values should relate to those same performance parameters" *(14)*.

Huber presented an equation that expresses this principle in mathematical form *(17)*. The equation was derived from flow–impedance relationships for a traffic stream consisting of basic vehicles (passenger cars) only and for a mixed traffic stream with a proportion of trucks, p, and a proportion of passenger cars $(1 - p)$. The equation, which expresses the PCE value as a function of the basic and mixed flow rates, q_B and q_M, that are equivalent in terms of the measure of impedance used to define LOS, is stated as follows:

$$\text{PCE} = (1/p)\ [(q_B/q_M) - 1] + 1 \quad (1)$$

Effect of Trucks

The adverse effect of trucks on traffic-stream performance can be attributed to three factors:

1. Trucks are larger than passenger cars,
2. Trucks have operating capabilities that are inferior to those of passenger cars, and
3. Trucks have a physical impact on nearby vehicles and a psychological impact on the drivers of those vehicles.

The first two are the factors that have traditionally been considered *(1, 18)*. Krammes suggested that truck-related problems—such as aerodynamic disturbances, splash and spray, sign blockage, offtracking, and the underride hazard—may also contribute to capacity reductions because of their effect on how nearby vehicles use the roadway *(19)*.

For capacity analysis purposes, roadways are divided into several basic types: freeways (with basic, ramp, and weaving sections), rural highways (multilane or two lane), and urban streets (signalized or unsignalized intersections, arterial streets). The relative importance of the three previously described factors on the overall effect of trucks differs among the roadway types. For example, the impact of the inferior operating capabilities of trucks is more severe on two-lane rural highways than on multilane freeways, which provide more passing opportunities. The relative importance of each factor also depends on roadway characteristics, such as geometry and configuration. For example, on sustained upgrades the impact of the inferior operating capabilities of trucks is "extremely deleterious" *(18)*; however, on level terrain, there is little difference between the speeds that passenger cars and trucks maintain *(1, 10, 12)*. Furthermore, the effect of trucks on nearby vehicles may be more important in certain roadway configurations—such as ramps or weaving sections, where lane changes are frequent—than in others—such as level, basic freeway sections, where fewer lane changes occur.

Therefore, the formulation to estimate PCEs for a particular roadway type should be expressed in terms of variables that reflect the combination of factors contributing to the overall effect of trucks on the quality of service provided by that roadway type.

APPROACHES TO ESTIMATING PCEs FOR LEVEL FREEWAY SEGMENTS

This section includes a historical review of PCE values recommended for trucks on level freeway segments and an evaluation of the merits of three approaches to estimating PCEs.

Historical Review

The 1950 *HCM* introduced the estimate that, on multilane highways in level terrain, trucks have the same effect as two passenger cars *(20)*. The *HCM* intimates that this estimate was based on the number of passings of trucks by passenger cars compared with the number of passings of passenger cars by passenger cars.

The 1965 *HCM* formally introduced both the LOS concept and the term PCE *(1)*. LOS was defined in terms of two parameters: operating speed and volume-to-capacity ratio. However, the PCE value of 2.0 for trucks on freeways in level terrain was a carry over from the 1950 *HCM (20)*.

Roess, McShane, and Pignataro recommended a revised approach to freeway LOS, using average running speed and density as the defining parameters *(21)*; this revised approach

was incorporated into *Transportation Research Circular 212 (18)*, which placed emphasis on density as the "primary measure of effectiveness" *(22)*. However, *Circular 212* continued to use a PCE value of 2.0 for trucks in level terrain *(18)*.

The 1985 *HCM* continues to define LOS in terms of density and average running speed but has revised downward the PCE for level terrain to a value of 1.7 *(15)*. Roess and Messer explain the reason for this revision *(14)*:

> The Institute for Research study *[13]* does, however, suggest that the PCE values currently used in the 1965 Highway Capacity Manual and in Circular 212 are higher than necessary. For example, the maximum PCE value of 2.0 applies only to tractor-trailers under the highest volume conditions. Maximum PCE values for single-unit trucks are 1.5 or 1.6, depending on the number of axles. . . . On the basis of these results, slight reductions in the level terrain PCE values of Circular 212 appear to be in order.

Alternative Approaches

Roess and Messer identified three approaches to estimating PCEs *(14)*:

1. The constant volume-to-capacity ratio approach,
2. The equal-density approach, and
3. The spatial headway approach.

The applicability of these approaches to freeway facilities is discussed in the following paragraphs.

The constant volume-to-capacity approach was appropriate when LOS was defined in terms of volume-to-capacity ratios. However, it is not applicable to the current procedure, which defines LOS using density and average running speed. Traffic streams that are equivalent in terms of volume-to-capacity ratio do not necessarily have equal speeds or densities.

The principal advantage of the equal-density approach is that density is the primary parameter used to define LOS. PCE values have not been estimated with this approach so far, although Huber developed a formulation with equal total travel time, which is numerically equal to density, as the basis for equivalence *(17)*. He used the linear relationship between speed and density, which was postulated by Greenshields *(23)*, to derive the formulation. Huber demonstrated that mixed and basic traffic streams that have equal densities operate at different speeds. As a characteristic of an approach for estimating PCEs this is undesirable because speed is the secondary parameter for defining LOS. This characteristic is also inconsistent with the intent of using density as the primary parameter for defining LOS; according to Roess, density is used because "it quantifies the proximity to other vehicles, and is directly related to the freedom to maneuver within the traffic stream" *(22)*. Certainly, when operating on the same freeway segment, traffic streams that have different speeds must have different degrees of freedom to maneuver. These observations lead to the conclusion that the basis for equivalence should not be equal density, but rather densities that feel the same to the driver in terms of proximity to other vehicles and freedom to maneuver. But how can this basis be implemented?

The answer lies in the spatial headway approach. As Roess and Messer note, "Average spacing and density are related on a one-to-one basis, and spatial headway could be argued to be a surrogate (more easily measured) parameter for density" *(14)*. The headway approach uses actual measurements of the relative position maintained by drivers in the traffic stream under prevailing conditions. Such measurements, if obtained in appropriate situations and with proper experimental control, should reflect the position that a driver chooses to maintain with respect to other vehicles. Those spacings maintained by drivers in the proximity of trucks and those maintained by drivers in the proximity of passenger cars should be equivalent in terms of the driver's perception of proximity to other vehicles and freedom to maneuver. Therefore, a formulation that properly relates these spacings should represent the driver's perception of equivalent densities.

FORMULATION OF HEADWAY APPROACH TO EQUIVALENCY

The derivation of a formulation that estimates PCEs based on the driver's perception of equivalent densities is described, and how to obtain appropriate headway measurements for use in the formulation is discussed.

Derivation of Formulation

The formulation is derived by introducing appropriate headway measurements into Huber's equation for PCEs, which was stated in Equation 1. This equation can be expressed in terms of time headway by introducing the fundamental relationship between flow rate and average time headway:

$$q_i = (3{,}600 \text{ sec/hr}) / \bar{h}_i \quad (2)$$

where q_i is the flow rate of vehicles per hour for either a basic stream ($i = B$) or an equivalent mixed stream ($i = M$); and $\bar{h}_i$ is the mean time headway in seconds at that flow rate.

Substituting Equation 2 into Equation 1 and rearranging yields

$$PCE = (1/p)[(\bar{h}_M - \bar{h}_B) / \bar{h}_B] + 1 \quad (3)$$

IFR advocated the use of lagging time headway, which includes the length of a vehicle and the intervehicular spacing that precedes the vehicle, as shown in Figure 1 *(13)*. The results of a statistical analysis by Krammes suggest that intervehicular spacings are affected by the types of the vehicles that delimit the spacing *(19)*. Because the objective is to derive a formulation for PCEs based on the driver's perception of equivalent proximity and freedom to maneuver and because the types of both the vehicle of interest and the leading vehicle may influence this equivalence, the headways in Equation 3 should be expressed in terms of the mean lagging time headways for each combination of pairs of vehicle types that are found in the traffic stream. The headways for each combination are expressed as $\bar{h}_{i_{jk}}$, where j refers to the vehicle of interest type

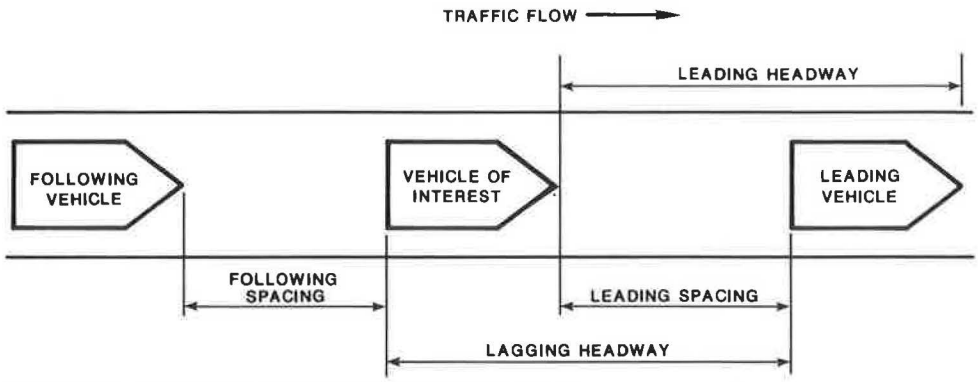


FIGURE 1 Schematic of vehicle headway and spacing measurements.

(either P for passenger car or T for truck) and k refers to the leading vehicle type.

Because the basic stream contains only passenger cars,

$$\bar{h}_B = \bar{h}_{B_{PP}} \tag{4}$$

In a mixed stream, four combinations of pairs of vehicle types occur. If the sequence of vehicle types in the mixed stream is random, the proportion of each combination in the traffic stream is the product of the proportion of each vehicle type. The assumption of randomness of sequencing of vehicle type in a mixed stream was examined by Krammes, whose results showed a slightly higher proportion of vehicle pairs of like type than would be predicted if sequencing were indeed random *(19)*. With this caveat in mind, the assumption of randomness of sequencing leads to the following expression:

$$\bar{h}_M = (1-p)^2\ \bar{h}_{M_{PP}} + p\,(1-p)\ \bar{h}_{M_{PT}} + p(1-p)\ \bar{h}_{M_{TP}} + p^2\ \bar{h}_{M_{TT}} \tag{5}$$

Substituting Equations 4 and 5 into Equation 3 yields

$$PCE = (1/p)\{[(1-p)^2\ \bar{h}_{M_{PP}} + p\,(1-p)\ \bar{h}_{M_{PT}} + p(1-p)\ \bar{h}_{M_{TP}} + p^2\ \bar{h}_{M_{TT}} - \bar{h}_{B_{PP}}] / \bar{h}_{B_{PP}}\} + 1 \tag{6}$$

Equation 6 could be simplified by assuming that $\bar{h}_{B_{PP}} = \bar{h}_{M_{PP}}$, which means that, on average, the lagging time headway a passenger car driver maintains when following another passenger car in a basic stream with flow rate q_B is the same as in a mixed stream at flow rate q_M and that the difference in flow rates and in mean lagging time headways for basic and mixed streams is due solely to the presence of trucks. This assumption implies that passenger car drivers in a mixed stream are affected only by trucks that are immediately preceding them. Research that analyzed the effect of following vehicle type on in-lane driver behavior *(19)* and research that examined the effect of the second vehicle ahead on car-following behavior *(24, 25)* support this implication.

By making this assumption, Equation 6 can be simplified to the formulation

$$PCE = [(1-p)\,(\bar{h}_{M_{PT}} + \bar{h}_{M_{TP}} - \bar{h}_{M_{PP}}) + p\ \bar{h}_{M_{TT}}] / \bar{h}_{M_{PP}} \tag{7}$$

This formulation has the advantage of using headway measurements from the mixed stream only. Estimates of PCEs for a site can be developed with data from that site only. Therefore, data would not be required from similar facilities that are used only by passenger cars: such facilities could be difficult to find. Also, problems of consistency, which could arise in using data from different facilities, would be avoided.

Krammes found that, after controlling for flow rate and speed, the effect of the leading vehicle type on the spacings maintained by combination trucks was significantly different from the effect on the spacings maintained by passenger cars *(19)*. Trucks maintained a significantly smaller spacing when traveling behind a leading truck than a leading passenger car (at a 95 percent confidence level), whereas passenger cars maintained slightly, but not significantly, larger spacings when traveling behind leading trucks than leading passenger cars. These findings apply to a data base that represented flow rates less than 1,300 vehicles per hour per lane. It may be reasonable to hypothesize that the effect of leading vehicle type on the spacings maintained by passenger cars would be significant at higher flow rates than were represented in the data base analyzed. Nonetheless, acceptance of these findings leads to the assumption that $\bar{h}_{M_{PT}} = \bar{h}_{M_{PP}}$, in which case Equation 7 would be reduced to

$$PCE = [(1-p)\ \bar{h}_{M_{TP}} + p\ \bar{h}_{M_{TT}}] / \bar{h}_{M_P} \tag{8}$$

where $\bar{h}_{M_P}$ refers to the mean lagging time headway for passenger cars, averaged across both leading vehicle types.

If it were further assumed that $\bar{h}_{M_{TP}} = \bar{h}_{M_{TT}}$, an assumption that the research by Krammes did not support *(19)*, then Equation 8 would be reduced to the following formulation

$$PCE = \bar{h}_{M_T} / \bar{h}_{M_P} \tag{9}$$

IFR used Equation 9 to estimate PCE values for use in a highway cost allocation study *(13)*. However, Equation 7 is recommended as the final formulation for use in highway capacity analysis because it accounts for the effect of leading vehicle type on the driver's perception of equivalent densities.

Although Equation 8 may be valid at low flow rates, Equation 7 would be equally valid at low flow rates and may be more accurate at higher flow rates.

Appropriate Headway Measurements

IFR computed overall means for passenger cars, $\bar{h}_{M_P}$, and trucks, $\bar{h}_{M_T}$, from data collected at 11 freeway sites in four urban areas for use in Equation 9 *(13)*. However, Equation 7 requires estimates of mean lagging time headways for the four combinations of pairs of passenger cars and trucks in a mixed traffic stream. A more sophisticated estimation procedure for estimating these headways, a procedure with features particularly suitable for highway capacity analysis, is described herein.

Headways should be measured while drivers are exhibiting steady-state, in-lane behavior. This implies, first, that drivers have maintained their lane placement and their position relative to other vehicles in the lane over some length of roadway and, second, that they have had the opportunity to adjust their speed and spacing relative to the leading vehicle. A sample of headways for vehicles exhibiting such behavior should reflect the spacings in the proximity of passenger cars and of trucks that are equivalent to the driver.

The data collected by IFR at six-lane, basic freeway segments on the Kingery Expressway in Chicago and on the La Porte Freeway in Houston were used in this analysis *(13)*. Drivers of a vehicle of interest were assumed to be exhibiting steady-state, in-lane behavior if they maintained the same lane placement and same position with respect to the leading and following vehicles for 300 ft before and after the point of measurement.

An analysis of covariance model was used to estimate the mean lagging time headways that are equivalent to the driver; the model has the following form:

$$\text{LNLTHD} = \text{INTERCEP}_{ijk} + B^1_{ijk}\text{LTYPE} + B^2_{ijk}\text{INVQ} + B^3_{ijk}\text{SPEED} + B^4_{ijk}\text{LSPEED} \quad (10)$$

where

LNLTHD = natural logarithm of lagging time headway (sec);
INTERCEP = parameter estimate for intercept;
i = site—Kingery or La Porte;
j = vehicle of interest type—passenger car or truck;
k = lane—1, 2, or 3;
B^1, B^2, B^3, B^4 = parameter estimates:
LTYPE = leading vehicle type—0 = passenger car, 1 = truck;
INVQ = [(3,600/flow rate in Lane k) − 6.00] (sec);
SPEED = speed of vehicle of interest − 55.0 (mph); and
LSPEED = speed of leading vehicle − speed of vehicle of interest (mph).

The dependent variable in the model is the natural logarithm of lagging time headway, LNLTHD. The logarithmic form is used because headway measurements have been found to fit a lognormal distribution *(26, 27)*. Separate equations are provided for each site, vehicle of interest type, and lane because Krammes has found that the leading intervehicular spacing maintained by a vehicle of interest is significantly affected by these variables and because it is unlikely that the effect is additive in nature *(19)*. Krammes also found that INVQ, SPEED, and LSPEED had a significant effect on intervehicular spacing and that the parameter estimates for LTYPE were significantly different in the equations for passenger cars and trucks *(19)*.

The overall R-square value for the model was 0.07, which reflects the tremendous variability in observed headways. The data support an observation by Breiman et al. that the mean and standard deviation of observed headways at a particular volume level are approximately equal *(28)*. Therefore, even though the variables in the model are significant, they explain only a small percentage in the tremendous variability in headways.

The data with which the model was calibrated represent a range of flow rates from approximately 400 to 1,300 vehicles per hour per lane. Therefore, to avoid extrapolating too far beyond the limits of the data, predicted values were estimated only for flow rates and speeds that approximate the upper limits of LOS A, B, and C. The flow rates (700, 1,100, and 1,550 passenger cars per hour per lane) and the speeds of the vehicle of interest (60, 57, and 54 mph) define the upper boundaries for LOS A, B, and C on basic freeway segments *(15)*. The relative speed of the leading vehicle was assumed to be zero because Krammes found that the difference between the speeds of passenger cars and trucks in a particular lane and at a particular volume level was generally less than 1 mph *(19)*. The predicted values for lagging time headway that correspond to these flow rates and speeds are presented in Tables 1 and 2 for the Kingery and La Porte sites, respectively.

Estimated PCE Values

Table 3 gives the estimates of PCE values for each lane and LOS at each site. These estimates were computed from Equation 7 by using the predicted values summarized in Tables 1 and 2 and the proportions of trucks for each lane and LOS. An estimate of the overall PCE value for each LOS, for all lanes combined, is also provided. This overall value is a weighted average of the value for each lane, weighted according to the distribution of trucks by lane at each LOS. This weighting scheme follows the approach recommended by Branston, who warned that PCE values that are based on a simple average of measurements for all lanes at a site may be inaccurate *(12)*.

The emphasis of this paper is on a theoretically based PCE formulation. The estimates in Table 3 are provided to demonstrate the approach. The PCE values fall within the range of values estimated by previous researchers *(1, 7, 13)*. The values estimated by IFR using Equation 9 are also included for comparison *(13)*. Because the values in Table 3 were based on limited data, especially at the lowest and highest flow rates, the actual values should not be considered precise.

TABLE 1 PREDICTED VALUES FOR LAGGING TIME HEADWAY AT KINGERY SITE (sec)

Lane	Vehicle of Interest Type[a]	Leading Vehicle Type[a]	Level of Service A	B	C
Right	P	P	3.89	2.62	1.99
		T	4.10	2.76	2.10
	T	P	5.12	4.35	3.90
		T	3.92	3.33	2.99
Center	P	P	3.80	2.34	1.71
		T	3.67	2.26	1.65
	T	P	3.72	2.73	2.20
		T	3.10	2.27	1.83
Median	P	P	2.54	1.73	1.31
		T	3.02	2.05	1.55
	T	P	4.23	3.37	3.13
		T	1.37	1.09	1.01

[a]P = passenger car and T = truck.

TABLE 2 PREDICTED VALUES FOR LAGGING TIME HEADWAY AT LA PORTE SITE (sec)

Lane	Vehicle of Interest Type[a]	Leading Vehicle Type[a]	Level of Service A	B	C
Right	P	P	3.65	2.91	2.48
		T	4.13	3.29	2.81
	T	P	4.91	4.32	3.92
		T	5.01	4.41	4.00
Center	P	P	3.24	2.37	1.92
		T	3.51	2.56	2.08
	T	P	4.10	3.53	3.28
		T	3.21	2.29	2.75
Median	P	P	2.76	1.97	1.54
		T	3.21	2.29	1.79
	T	P	3.64	3.52	3.39
		T	3.13	3.02	2.92

[a]P = passenger car and T = truck.

TABLE 3 ESTIMATES OF PCE VALUES FOR COMBINATION TRUCKS ON LEVEL FREEWAY SEGMENTS

Lane	Level of Service A	B	C
Kingery Site			
Right	1.2	1.6	2.0
Center	0.9	1.1	1.2
Median	1.8	2.1	2.6
All	1.0	1.2	1.2
LaPorte Site			
Right	1.5	1.6	1.7
Center	1.3	1.5	1.8
Median	1.5	1.9	2.3
All	1.4	1.6	1.8
Institute for Research Values			
All	1.1	1.2	1.4

However, three characteristics of the PCE values are interesting to note. First, the values for the two sites differ. Second, at a particular site, the values for each lane differ. Third, the values increase from LOS A to LOS C. Unfortunately, the statistical significance of these differences cannot be tested because the complexity of the PCE formulation makes the computation of confidence intervals intractable.

The difference between the PCE values for the two sites, lower values at the Kingery site than at the La Porte site, may reflect the differences in the percentages of trucks and in the truck management strategies at the two sites. At the La Porte site, the traffic stream included 10 percent trucks, whereas at the Kingery site there were 28 percent trucks. At the La Porte site, trucks were permitted in all lanes, whereas at the Kingery site trucks were prohibited from using the median lane. The truck management strategy at the Kingery site resulted in high percentages of trucks in the center lane (approximately 47 percent).

The PCE values for the center lane of the Kingery site are particularly interesting; these values are much lower than those for the other lanes at either site.The value of 0.9 at LOS A in the center lane of the Kingery site indicates that the mean lagging time headways for vehicle pairs including trucks are smaller than for pairs consisting of two passenger cars. Because trucks are larger than passenger cars, the reason for the smaller headways for trucks is that trucks maintained smaller leading intervehicular spacings than passenger cars. The resulting PCE values, which are considerably smaller for the Kingery site, overall and for the center lane in particular, suggest that the truck management strategy may be an effective way to minimize the adverse effect of trucks on freeway capacity.

The question of how PCE values vary with flow rate has been the subject of debate. The approach that this research recommends incorporates flow rate explicitly into the estimation procedure by including flow rate as an independent variable in the analysis of covariance model that estimates the headway measurements used to compute PCEs.

The proposed formulation estimates PCE values that increase with flow rate. IFR *(13)* and Cunagin and Messer *(7)* also found that PCE values increase with flow rate on level urban freeways and on level, four-lane rural highways, respectively. Huber in his author's closure states a preference for PCE values that increase with flow rate because "as the flow rate increases, the opportunity for interaction between basic vehicles and trucks is increased with a subsequent increase in PCE values" (*17*, p.69).

Both Roess and Messer *(14)* and St. John [in his discussion of Huber (*17*, pp. 68–69)] advocate PCE values that do not increase with flow rate because a constant value would simplify calibration of the values as well as computations with the values. St. John also observes that "constant PCE implies fundamental relationships that do not change in form between car-only and mixed flows" (*17*, pp 68–69). He cites results from a microscopic model of multilane flow, which imply that PCE values "would be essentially constant" *(29)*. Roess and Messer *(14)* also refer to these and other related results: "none of the studies looking at PCEs on specific grades showed significant variation with volume."

The responses to these arguments are as follows:

1. The available evidence does not clearly indicate that speed-flow relationships for car-only and mixed flows have the same form—St. John states that "there is evidence both supporting and conflicting with the idea that PCE is constant over flow rate" (*17*, pp. 68–69).
2. The data bases that have been used to estimate PCEs contained little or no data at high flow rates and, therefore, have not provided reliable estimates of PCEs over the entire range of flow rates.
3. The characteristics of PCE values for specific grades are not necessarily the same as those for level terrain.

The last point reinforces the desirability of the proposed theoretical basis for PCEs, which emphasizes that PCEs for a particular roadway type should reflect the effects of trucks on that roadway type and which provides a framework to account for the differences between PCEs for each roadway type. Although the current research suggests that PCEs increase with flow rate, it does not represent the final answer. St. John's conclusion appears appropriate (*17*, p. 69):

> I suggest that more attention be directed to the fundamental concepts of equivalence . . . . Also final decisions should be based on extensive field data or results from comprehensive models.

SUMMARY AND CONCLUSIONS

The needs for a commonly accepted definition of equivalence and for a clearly defined theoretical basis on which the concept of passenger car equivalency can be applied to any roadway type were addressed in this paper. Two principles were defined as components of the theoretical basis for estimating PCEs for capacity analysis:

1. The basis for equivalence should be the parameters used to define LOS for the roadway type in question.
2. The PCE formulation should be expressed in terms of variables that reflect the relative importance of the three factors that contribute to the effect of trucks on that roadway type.

Traditionally, the effect of trucks on highway capacity has been attributed to two factors: *(a)* trucks are larger than passenger cars, and *(b)* trucks have operating capabilities that are inferior to those of cars.

However, a third factor should also be considered: trucks have physical impacts on nearby vehicles and psychological impacts on the drivers of those vehicles that also contribute to reductions in capacity. It should be emphasized that research on the significance of this third factor has considered only the effect of leading and following trucks on the in-lane behavior of vehicles; no research has been performed on the effect of trucks on the lane-changing behavior of vehicles or on vehicles in adjacent lanes.

Huber's general equation for PCEs, which expresses the first principle in mathematical form, was used as the starting point for the derivation of a PCE formulation for level freeway segments that was expressed in terms of headways. The PCE value of 1.7, used in the 1985 *HCM* for trucks on level freeway segments *(15)*, was influenced by estimates made by IFR, which used a headway-based PCE formulation *(13)*. The assumptions inherent in IFR's formulation were identified. A more sophisticated formulation and estimation procedure was also discussed.

This more sophisticated headway-based approach has three advantages:

1. It accounts for the effect of leading trucks on the intervehicular spacings maintained by a vehicle of interest.
2. The percentage of trucks in the traffic stream, which has an important effect on PCE values, is included as a variable in the model.
3. The estimation procedure allows PCE values to be estimated for specific speeds and flow rates, enabling the effect of these variables to be considered explicitly.

ACKNOWLEDGMENT

The research described herein was conducted at the Pennsylvania Transportation Institute for Dr. Krammes's doctoral dissertation in civil engineering at the Pennsylvania State University. The contributions of Dr. Walter P. Kilareski, dissertation advisor, are gratefully acknowledged.

REFERENCES

1. *Special Report 87. Highway Capacity Manual—1965*, Highway Research Board, Washington, D.C., 1966, 411 pp.
2. D. W. Gwynn. "Truck Equivalency." *Traffic Quarterly*, Vol. 22, No. 2, 1968, pp. 225–236.
3. E. F. Reilly and J. Seifert. "Truck Equivalency." In *Highway Research Record 289*, HRB, National Research Council, Washington, D.C., 1969, pp. 25–37.
4. A. D. St. John. "Nonlinear Truck Factor for Two-Lane Highways." In *Transportation Research Record 615*, TRB, National Research Council, Washington, D.C., 1976, pp. 49–53.
5. A. Werner and J. F. Morrall. "Passenger Car Equivalencies of Trucks, Buses, and Recreational Vehicles for Two-Lane Rural Highways." In *Transportation Research Record 615*, TRB, National Research Council, Washington, D.C., 1976, pp. 10–16.
6. J. Craus, A. Polus, and I. Grinberg. "A Revised Method for the Determination of Passenger Car Equivalencies." *Transportation Research*, Vol. 14A, No. 4, 1980, pp. 241–246.
7. W. D. Cunagin and C. J. Messer. *Passenger Car Equivalents for Rural Highways*. Report FHWA/RD-82/132. H. G. Whyte Associates, Inc., Gary, Ind.; and Texas A&M University, College Station, 1982.
8. W. D. Cunagin and C. J. Messer. "Passenger-Car Equivalents for Rural Highways." In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp. 61–68.
9. M. Van Aerde and S. Yagar. "Capacity, Speed, and Platooning Vehicle Equivalents for Highways." In *Transportation Research Record 971*, TRB, National Research Council, Washington, D.C., 1984, pp. 58–65.
10. E. M. Linzer, R. P. Roess, and W. R. McShane. "Effect of Trucks, Buses, and Recreational Vehicles on Freeway Capacity and Service Volume." In *Transportation Research Record 699*, TRB, National Research Council, Washington, D.C., 1979, pp. 17–26.
11. W. Berman and R. C. Loutzenheiser. *Study of Traffic Flow on a Restricted Facility, Report III-2, A Methodology to Measure the Influence of Trucks on the Flow of Traffic*. Report FHWA-MD-R-77-7. University of Maryland, College Park, 1976.
12. D. Branston. "Some Factors Affecting the Capacity of a Motor-

way." In *Traffic Engineering and Control*, Vol. 18, No. 6, 1977, pp. 304–307.
13. E. L. Seguin, K. W. Crowley, and W. D. Zweig. *Urban Freeway Truck Characteristics*, Vol. I: *Passenger Car Equivalents*. Report FHWA/RD-81/156. Institute for Research, State College, Pa., 1981.
14. R. P. Roess and C. J. Messer. "Passenger Car Equivalents for Uninterrupted Flow: Revision of the *Circular 212* Values." In *Transportation Research Record 971*, TRB, National Research Council, Washington, D.C., 1984, pp. 7–13.
15. *Special Report 209*: *Highway Capacity Manual*, TRB, National Research Council, Washington, D.C., 1985.
16. R. P. Roess. "Highway Capacity Manual Revisited." *Civil Engineering*, Vol. 54, No. 11, 1984, pp. 69–71.
17. M. J. Huber. "Estimation of Passenger-Car Equivalents of Trucks in Traffic Stream." In *Transportation Research Record 869*, TRB, National Research Council, Washington, D.C., 1982, pp. 60–70.
18. "Interim Materials on Highway Capacity." *Transportation Research Circular 212*, TRB, National Research Council, Washington, D.C., 1980, 276 pp.
19. R. A. Krammes. *Effect of Trucks on the Capacity of Level, Basic Freeway Segments*. Ph. D. dissertation. Pennsylvania State University, University Park, Pa., 1985.
20. Highway Research Board. *Highway Capacity Manual*. U. S. Government Printing Office, Washington, D.C., 1950.
21. R. P. Roess, W. R. McShane, and L. J. Pignataro. "Freeway Level of Service: A Revised Approach." In *Transportation Research Record 699*, TRB, National Research Council, Washington, D.C., 1979.
22. R. P. Roess. "Level of Service Concepts: Development, Philosophies, and Implications." In *Transportation Research Record 971*, TRB, National Research Council, Washington, D.C., 1984, pp. 1–16.
23. B. D. Greenshields. "A Study of Traffic Capacity." *Proc., Highway Research Board*, Vol. 14, 1934, pp. 448–477.
24. R. Herman and R. W. Rothery. "Car Following and Steady State Theory." *Proc., Second International Symposium on the Theory of Road Traffic Flow, London, 1963*; J. Almond (ed.), Organization for Economic Cooperation and Development, Paris, 1965, pp. 1–11.
25. P. Fox and F. G. Lehman. *Safety in Car Following—A Computer Simulation*. Newark College of Engineering, Newark, N.J., 1967.
26. I. Greenberg. "The Log-Normal Distribution of Headways." *Australian Road Research*, Vol. 2, No. 7, 1966, pp. 14–18.
27. J. E. Tolle. "The Lognormal Headway Distribution Model." *Traffic Engineering and Control*, Vol. 13, No. 1, 1971, pp. 22–24.
28. L. Breiman, W. Lawrence, R. Goodwin, and B. Bailey. "The Statistical Properties of Freeway Traffic." *Transportation Research*, Vol. 11, No. 4, 1977, pp. 221–228.
29. A. D. St. John, D. Kobett, W. Glauz, D. Sommerville, and D. Harwood. *Freeway Design and Control Strategies as Affected by Trucks and Traffic Regulations*. Report FHWA/RD-75/42. Midwest Research Institute, Kansas City, Mo., 1975.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Abridgment

Transitions in the Speed-Flow Relationship

MARGOT A. GUNTER AND FRED L. HALL

Transitions between regimes in the speed-flow relationship were the topic of a recent paper by the authors (*Transportation Research Record 1005*, 1985). Subsequently, the authors have found that a different analytic technique provides better insights into this problem. This approach involves the use of daily time-traced and overlaid time-traced plots. The result of the re-analysis is that four previous conclusions are confirmed and one is not. Several questions that remained after the first paper about the shape of the relationship and its transitions are resolved, although why operation in shoulder lanes is different from that in the nonshoulder lanes is still unclear.

In this paper the focus is on the nature of transitions between the congested and uncongested regimes in freeway traffic flow, particularly in the speed-flow relationship. In a previous paper, the authors used an event-based averaging procedure to gain insights into the process involved in the breakdown and recovery of traffic flow *(1)*. Since that time, the authors have identified a better means of analysis to resolve the problem. Application of that technique has suggested that the event-based averaging was somewhat misleading and that perhaps the transitions were being incorrectly defined. One of the conclusions in the previous paper is therefore invalid, and several of the questions raised in it can now be resolved.

The analytic technique consists of examining a time-connected plot of speed and flow for each individual day. Preparation of such plots is not new. Ceder, for example, used them in his dissertation *(2)*. Using them to focus on transitions, as demonstrated in the next section, is perhaps a bit novel. The main discussion, contained in the third section, is based on overlaid time-connected plots for many days. The data utilized here are described by Allen et al. *(1)* and Hall et al. *(3)*, and in a paper by Hall and Gunter elsewhere in this Record.

EXAMPLES OF ANALYSIS

The analytic technique can best be described by displaying examples of the plots for a single lane at a specific location. Details of decisions made to arrive at these particular plots are discussed by Hall et al. *(3)* and in a paper by Hall and Gunter elsewhere in this Record. These examples help to demonstrate the variety of information that can be obtained by using this approach. Station 4, median lane is used because it *(a)* is geometrically ideal, and *(b)* experiences a wide range of flows. The station is about 4 km and three entrance ramps upstream of the primary bottleneck, and therefore regularly experiences transitions between regimes.

The flow-occupancy trace was found to be useful in this analysis because of the more easily identified critical operation (e.g., Figure 1a compared with Figure 1b), and because some 5-min intervals are identified to be transitional only by examining occupancy (e.g., the point in the middle of the return to uncongested flow in Figure 1b). From the trace of any individual day, it is possible to determine some information on the nature of any transitions (speed of occurrence, change in flow). Examination of the data is greatly aided by an overlaid time-traced plot (Figure 2), which provides a framework for the underlying relationship for all of the daily plots. Overall trends in the transitions can also be observed from the overlaid plot, such as the changes in flow that may be associated with the jumps, and the flow ranges at which they occur.

DISCUSSION OF ANALYSIS

The new approach is substantially different from the event-based averaging in the authors' previous paper *(1)*. Although four of the earlier conclusions are supported, one is altered. The five conclusions in the previous paper are summarized briefly as follows:

1. Transitions toward both regimes occur fairly rapidly.
2. Recovery to the uncongested regime occurs at almost constant volume.
3. Different locations along a highway require different speed-flow relationships to describe their operation.
4. The reduction of flow and low starting speed associated with the line representing transitions from upper to lower branch operation are not easily explained.
5. Speed-flow relationships differ according to lane.

In addition, several questions were left unanswered, as follows:

1. The exact shape of the underlying relationship near capacity is not clear.
2. The values of parameters such as critical speed and capacity flow were not ascertainable.
3. Operation in the shoulder lanes was very different from that in the nonshoulder lanes, and could not be explained.

Conclusion 4 is altered by the re-analysis: the line representing transitions to the congested regime can now be explained. The explanation is simply that the line shown to represent the average of transitions was incorrect. The reason that it was incorrect is that transitions from uncongested to congested

M. A. Gunter, IBI Group, 240 Richmond St. West, Toronto, Ontario M5V 1W1, Canada; formerly with McMaster University. F. L. Hall, Department of Civil Engineering, McMaster University, Hamilton, Ontario L8S, 4L7, Canada.

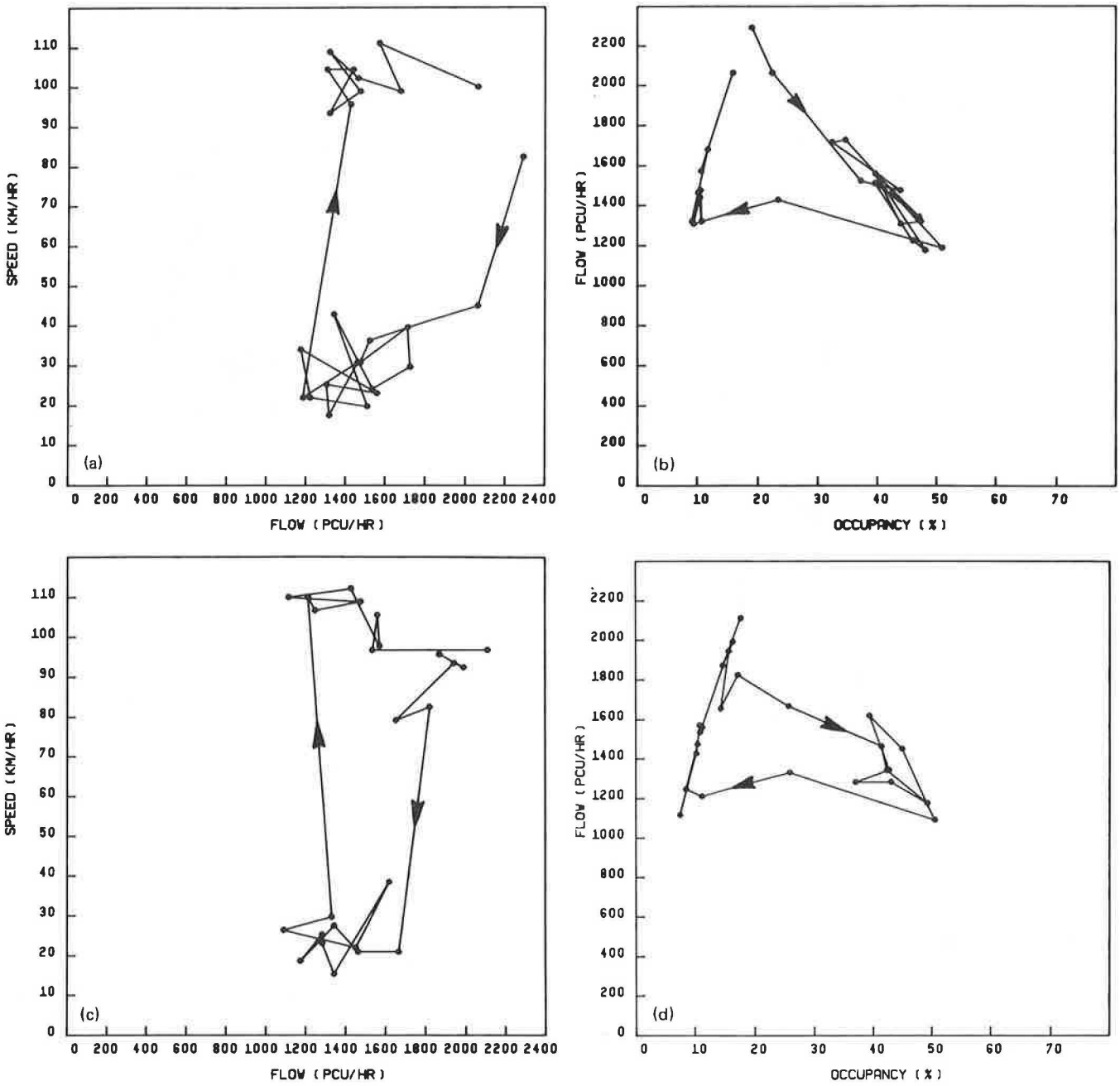


FIGURE 1 Examples of daily speed-flow (a and c) and flow-occupancy (b and d) time-traced plots for two days.

operation (such as in Figure 1c, with a large change in speed and occupancy at about the same flow) were being averaged with operation in the congested regime (Figure 1a) because the criterion to identify a transition was a speed change of greater than 15 km/hr. In other words, the daily time-traced plots indicated that the speed-change criterion could be met within the large fluctuations in the congested regime, such that events were identified that were not transitions. The other problem in the definition of events was that the equilibrium or nontransitional relationship had not been identified.

The other four conclusions made in the previous paper are supported by the current analysis. Rapid changes in speed and occupancy are still observed during both the breakdown to the congested regime and the recovery to the uncongested regime. These generally occur within a 10-min period, but on a few days take longer. Different speed-flow relationships are still observed for different lanes, as subsequent analysis of Station 4 shoulder lane indicates. Finally, the transition from congested to uncongested operation still appears to take place at approximately constant volume, at least in the median and middle lanes. This can be observed from both Figures 1 and 2, in which slight variations of flow rate certainly occur, but the average trend is a zero change. The slight variations can be easily explained by changes in mainline or ramp demand over the short transition period.

Two of the questions raised in the previous paper are answered by the re-analysis because the transition from uncongested to congested operation in a median or middle lane is

characterized by a speed and occupancy change at almost constant flow rate. Some confusion in the authors' previous discussion was caused because of the misleading averages, primarily as to the implications for the shape of the underlying relationship. It has now been determined that the speed at which maximum flow is observed is between 85 and 90 km/hr for the Station 4 median lane and above 80 km/hr for the other six median and middle lanes examined. Because these speeds occur at very high flow rates (2,350 pcu/hr for the Station 4 median lane), it can be assumed that operation is approaching capacity and that the critical speed is only slightly lower. The

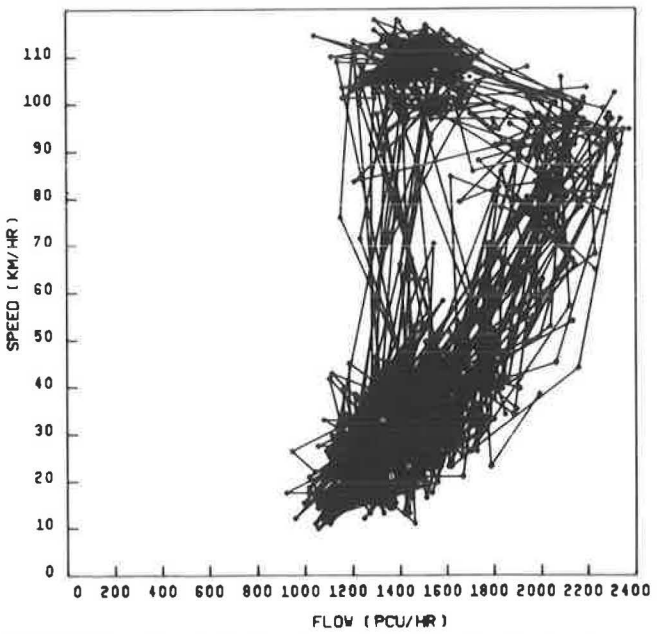


FIGURE 2 Overlaid time-traced plot for the Station 4 median lane.

resultant critical speed will therefore be substantially higher than the 50 km/hr suggested by the new *Highway Capacity Manual* (HCM) *(4)*.

One question remains unresolved: explanation of the operation in the shoulder lanes, which is obviously different from that in the nonshoulder lanes and remains difficult to explain. The major difference is in the shape of the observed relationship. Even well upstream of any entrance ramp four differences can be observed (Figure 3). First, uncongested speeds are much lower. Second, speeds change more quickly with flow in the uncongested regime than they do in other lanes. Third, maximum flows are much lower, and are observed in the congested regime.

The fourth striking difference is that the transitions identified toward both regimes do not occur at constant flow rates. Recoveries tend to occur with a decrease of 200 to 400 pcu/hr in flow, and if breakdowns occur they do so with a similar increase in flow. Two possibilities exist for the change from uncongested to congested flow: *(a)* transitions occur with an increase of flow between regimes on an underlying relationship similar to that in the nonshoulder lanes; or *(b)* operation moves along a different equilibrium relationship [which resembles the

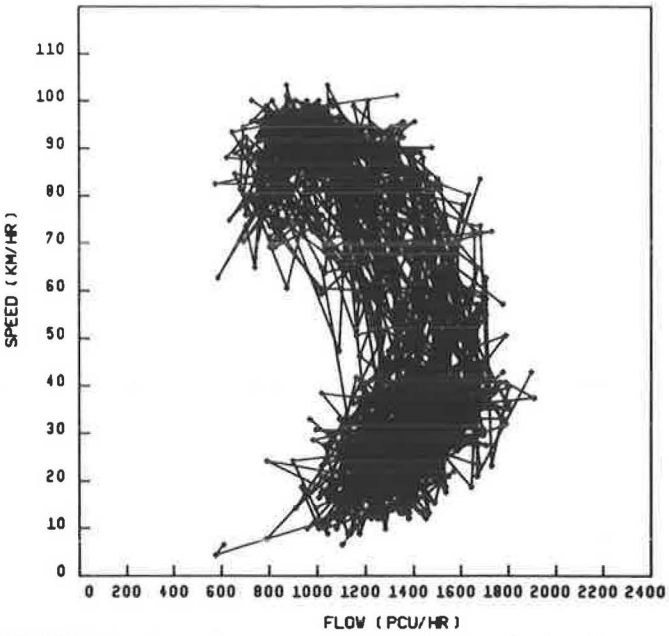


FIGURE 3 Overlaid time-traced plot for the Station 4 shoulder lane.

curve in the 1965 HCM *(5)*]. If it is on an HCM-type curve, then capacity would be 1,600 to 1,800 pcu/hr, depending on proximity to entrance ramps.

Thus, the question of why operation on the shoulder lanes is different from that in the nonshoulder lanes remains unanswered. An alternative method for identifying equilibrium and nonequilibrium flows is required [perhaps that used by Payne *(6)*].

CONCLUSION

The primary conclusion of this paper is that the use of daily time-traced and overlaid time-traced plots is the preferred method when analyzing transitions in the speed-flow relationship. Use of this analytic technique has demonstrated that the event-based averaging used in a previous paper on the subject could incorrectly identify the transitions. Re-analysis of the authors' data with the new technique changed one conclusion of the original paper and answered several questions raised in the original discussion.

ACKNOWLEDGMENTS

The authors would like to thank the Ontario Ministry of Transportation and Communications for providing the data for this analysis, and Geddes Mahabir, who several years ago made the data manageable on McMaster University's computer facilities. In addition, the authors should acknowledge the helpful and thought-provoking comments of two reviewers of our previous paper. Financial support was received from the Natural Sciences and Engineering Research Council through their Undergraduate Summer Research Award programme.

REFERENCES

1. B. L. Allen, F. L. Hall, and M. A. Gunter. Another Look at Identifying Speed-Flow Relationships on Freeways. In *Transportation Research Record 1005*, TRB, National Research Council, Washington, D.C., 1985, pp. 54–64.
2. A. Ceder. *Investigation of Two-Regime Traffic Flow Models at the Micro- and Macro-scopic Levels*. Ph.D. dissertation. University of California, Berkeley, Jan. 1975.
3. F. L. Hall, B. L. Allen, and M. A. Gunter. Empirical Analysis of Freeway Flow-Density Relationships. *Transportation Research*, Vol. 20A, Elmsford, N.Y., 1986, pp. 197–210.
4. *Special Report 209: Highway Capacity Manual*, TRB, National Research Council, Washington, D.C., 1985.
5. *Special Report 87: 1965 Highway Capacity Manual*, HRB, National Research Council, Washington, D.C., 1966. 411 pp.
6. H. J. Payne. Discontinuity in Equilibrium Freeway Traffic Flow. In *Transportation Research Record 971*, TRB, National Research Council, Washington, D.C., 1984, pp. 140–146.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Abridgment

FREESIM: A Microscopic Simulation Model of Freeway Lane Closures

AJAY K. RATHI AND ZOLTAN A. NEMETH

Development of a model to simulate traffic operations at freeway lane closures is described. The model logic is based on a rational description of the behavior of the drivers in a freeway lane-closure situation. The simulation program is written in SIMSCRIPT II.5 programming language. An application of the model is given with evaluation of potential safety impacts of reduced speed zones in freeway lane closures at different levels of assumed driver compliance.

With the Interstate system nearly complete, the emphasis has now shifted toward continued maintenance of this freeway network. Resurfacing, upgrading, and other corrective measures are required to maintain the original design standards and to eliminate previously unrecognized safety hazards. Construction and maintenance work activities requiring temporary closure of a freeway lane represent a frequently encountered and potentially hazardous situation. A study of road-under-repair accidents in Virginia found, for example, that of 426 accidents (for which the information on traffic control characteristics was available), 47.9 percent occurred at lane closures *(1)*.

A. K. Rathi, KLD Associates, Inc., 300 Broadway, Huntington Station, N.Y. 11746. Z. A. Nemeth, The Ohio State University, Columbus, Ohio 43210.

Freeway lane closures require properly developed traffic control plans to minimize the disturbance to the traffic flow and provide for the safety of both drivers and the working crew. Drivers approaching a work zone in the closed lane must receive and understand the information that they need to change lanes and merge into the traffic in the open lane. Although this in itself does not appear to be an unusually demanding driving task, problems still appear to develop, resulting in rear-end collisions, sideswipes, and single-vehicle fixed-object accidents *(2)*.

Improving traffic control systems requires comprehensive information on the relationship between control strategies and the quality of traffic flow (e.g., delay, travel time). Computer simulation provides an excellent basis for evaluating a wide sprectrum of traffic management schemes within the framework of controlled experiments. In this paper, development, verification, validation, and application of a microscopic simulation model (FREESIM) of traffic operations at freeway lane closures is described.

FREESIM SIMULATION MODEL

FREESIM is a microscopic, stochastic simulation model; vehicles are represented individually and their detailed, time-vary-

ing behavior is simulated. The model logic is based on a rational description of the behavior of the drivers in a lane-closure situation.

Vehicles are generated randomly from a shifted negative exponential distribution of arrival time headways, and are advanced in the system using a classical car-following approach; that is, each vehicle attempts to advance at its desired speed while maintaining a safe following distance from the vehicle ahead.

The lane change from the closed lane is described by a stimulus-response approach *(3)*. The basic behavioral assumption of the stimulus-response approach is that each traffic control device (or the view of the lane closure itself) can be considered as a stimulus and each will induce a different proportion of drivers to change lanes.

The vehicles are processed through the conventional gap-acceptance procedure to determine whether the lane change is possible.

The stimulus-response probability distribution was calibrated based on a survey of 229 drivers conducted at several freeway lane-closure sites *(4)*. The information processing time and the response initiation time for traffic control devices used in freeway lane closures were calibrated by using data from McGee et al. *(5)*.

The representation of the traffic control devices in the simulation program includes the advance warning signs recommended in the Manual on Uniform Traffic Control Devices (MUTCD) and a flashing arrow board for single-lane closure in multilane freeways *(6)*. FREESIM, as an option, can simulate the effect of a reduced-speed zoning by specifying the parameters (i.e., legibility, information processing time, and response initiation time) for a speed-limit sign and the assumed proportion of the drivers complying with it.

FREESIM can simulate single-lane closure of a freeway section with two or three unidirectional lanes of any reasonable length. For each lane, up to five data collection points can be simulated at user-specific locations to collect data on the ongoing behavior of the vehicles in the system, namely, speed and headway distributions. An additional data collection point is provided (by default) at 500 ft upstream of the system exit point.

SIMULATION INPUT

The input data (all free format) required for a simulation run include four types:

1. Simulation run parameters: simulation time, warm-up time, time interval between vehicle trajectories data collection, time interval between status listings, and random number seeds;
2. Roadway parameters: length of freeway section, location of the taper and of the headway, and spot speed data collection points;
3. Traffic parameters: approach volume, proportion of trucks, mean speed factors, and standard deviation of mean speed;
4. Traffic control device parameters: preference, legibility distance, type and speed limit; and
5. Vehicle and driver characteristics: length and maximum emergency deceleration rate of various vehicle types and mean brake reaction time of the drivers.

OUTPUT DESCRIPTION

The standard output of the program includes a title page and a list of all relevant input parameters. The simulation results are printed in the following general classification:

1. Simulated data collection points information: distributions and summary statistics of spot speeds and time headways at user-specified locations;
2. Summary statistics on measures of performance: travel time, delay, and speed gradient;
3. Volume-throughput data: throughput, input, and output volumes;
4. Lane-changing data: frequency distribution and statistics of lane-change initiation points and gap acceptance; and
5. Deceleration: number of uncomfortable and emergency decelerations.

PROGRAM DESCRIPTION

FREESIM was developed using SIMSCRIPT II.5 programming language, developed by CACI *(7)*. SIMSCRIPT II.5 is a powerful simulation programming language, both for discrete-event and continuous simulation. The SIMSCRIPT programs are composed of free-form, English-like statements; hence, the SIMSCRIPT II.5 source program becomes a useful part of the model documentation.

FREESIM is implemented on an AMDAHL 470/V8 system, which is an IBM-compatible computer. For a 6,000-ft freeway section, the average ratio of real time to central-processing-unit time varied from 70:1 to 40:1 as the two-lane approach volume varied from 1,200 to 1,800 vehicles per hour.

VERIFICATION AND VALIDATION

Verification of the simulation program was performed by operational testing of the car-following and lane-changing algorithms and sensitivity analysis of measures of effectiveness to the exogeneous (input) variables.

In the operational testing of the car-following algorithm, velocity responses of the vehicles to artificially induced speed disturbances and a blockage were observed. The tests indicated that the car-following algorithm shows all the desired characteristics: realism, stability, and reasonable oscillatory following behavior.

In the operational testing of the lane-changing component, the lane-changing behavior was observed for a range of values of traffic volumes in the closed lane of a two-lane freeway section. This testing emphasized the

- Compatibility of the lane-changing algorithm with the car-following algorithm, and

- Performance of the lane-changing algorithm at high levels of traffic volume.

Sensitivity analysis indicated that the measures of effectiveness (e.g., mean speeds and time headways at various locations in the simulated freeway section, merging proportions) were sensitive to the changes in input parameters for the lane mean speeds and stimulus-response probability distribution. The changes (up to 20 percent) in other input variables (e.g., information processing time, proportion of trucks, mean brake reaction time, and maximum emergency deceleration rate) did not have any appreciable effect on the simulation output. The implication of this analysis is that a strong data base is necessary for the calibration of the lane mean speeds and the stimulus-response probability distribution.

The simulation program was validated by comparing simulated time headway, speed, and merging distributions with four sets of actual observations obtained from three different rural freeway lane-closure sites *(4)*.

Compared with data from the field studies, the simulation model reproduced headway and merging distance distributions accurately. The comparison of simulated versus observed speed distributions indicated reasonably good agreement.

Also, the simulation model outputs representing the macroscopic traffic flow parameters (i.e., speed, flow, and throughput) and lane-changing frequencies were compared with some well-known empirical data from the literature. The simulation output compared well with the data from the revised Highway Capacity Manual *(8)* and Greenshield's Model *(9)*. Similarly, the simulated lane-changing frequencies were compatible with the data from the Northwestern University lane-changing study and INTRAS simulated observations *(10)*.

MODEL APPLICATION

A practical application of the model is given with simulation experiments to evaluate potential safety impacts of reduced speed zones at freeway lane closures at different assumed levels of compliance. The introduction of reduced speed zones in freeway lane closures is a controversial and not well understood aspect of traffic control *(11)*. Although reduced speed implies greater safety, at least intuitively, it introduces a disturbance in the traffic flow that may have negative impact—that is, increased probability of shock-wave formation upstream of the construction zone. The problem can be critical at high approach volumes.

A fractional factorial design was developed for the analysis of three independent variables: speed limit, compliance with the speed limit, and two-lane approach volume. Compliance levels of 0.33, 0.66, and 1.00 were used for each of the two reduced speed limits implemented: 50 mph and 45 mph. Four levels of two-lane approach volumes (vehicles per hour) were used: 800, 1,200, 1,500, and 1,800.

Three safety-related measures of performance from FREESIM output were selected as the dependent variables: number of uncomfortable decelerations per hour, variance of speed, and proportion of headways less than 2 sec in open lane at beginning of transition zone. An uncomfortable deceleration is defined as one that exceeds by more than 2 ft/sec/sec the deceleration rate normally considered comfortable at a given speed. The variance of speed at the transition zone is generally considered as the proxy variable for potential rear-end accidents upstream. Similarly, the presence of short headways in a transition zone can be considered as increasing the likelihood of rear-end collisions in the approach area.

SIMULATION RESULTS

The results of simulation experiments for the approach volume level of 1,800 vehicles per hour are given in Table 1. The results presented in the table are based on the average of five replications. The simulation results for all volume levels are described in detail elsewhere *(12)*. These results are presented here for the sole purpose of demonstrating the application of the model.

TABLE 1 SIMULATION RESULTS

Speed Limit (mph)	Compliance Level	N^a	V^b	P^c
55		1,258.0	102.1	69.8
	0.33	1,218.0	97.1	68.2
50	0.66	1,218.0	107.5	67.2
	1.00	1,514.8	112.4	67.5
	0.33	981.6	83.1	67.9
45	0.66	1,178.0	81.6	68.0
	1.00	1,052.0	84.7	67.8

^aN = uncomfortable deceleration/hr.
^bV = variance of speed (mph^2).
^cP = proportion of headways less than 2 sec.

The introduction of the 45-mph speed zone reduced the number of incidents of uncomfortable decelerations at all levels of compliance. However, the 50-mph speed zone was ineffective. In summary, it required a 10-mph reduction in the desired speed of the simulated group of complying drivers to produce the desired impact on the traffic flow.

The results were similar with regard to the speed variance measured at the beginning of the taper. The impact on headway distribution, measured by the proportion of headways less than 2 sec, was small but consistent in all cases.

The impact of the introduction of speed zones at lane closures depends heavily on approach volumes. The data presented in Table 1 represent high-volume conditions. By generating speed data at the beginning of the taper area, it was possible to demonstrate that approach speeds were constrained by near-capacity conditions; consequently, the introduction of speed zones had little potential to improve conditions. The benefits were considerably different at lower volumes.

OTHER POTENTIAL APPLICATIONS

FREESIM can be used for a variety of practical applications, in particular, evaluation of traffic performance in freeway lane closures under different control schemes.

In the MUTCD *(6)* the recommended traffic control treatment for typical freeway lane closures is described. However,

accident experience at lane-closure locations indicates that the minimum standards applicable for typical situations may not be adequate for all situations. FREESIM can be conveniently used to evaluate the impact of alternative advance warning signs or novel experimental signs on traffic performance in freeway lane closures. A before-and-after study could be conducted to evaluate the changes in traffic performance resulting from different control schemes. In the same format, FREESIM can also be used to evaluate the effect of nonconformity to the MUTCD standards such as sign misplacement or omission, poor maintenance, and use of confusing message. In these applications, the recalibration of the stimulus-response probability distribution would be required, particularly because the output variables were found to be sensitive to changes in this distribution.

An interesting application of the simulation model would be to assess the influence of problem drivers—for example, inattentive drivers or risk takers—on the behavior of the system. The influence of target driver groups such as speeders and risk takers can be included by appropriately modifying the simulation input parameters of drivers' performance characteristics, for example, speed distribution and distribution of brake reaction time.

The simulation program can also be used in optimization applications, such as the determination of optimum length of a single-lane zone for different approach volumes.

REFERENCES

1. B. T. Hargroves and M. R. Martin. *Vehicle Accidents in Highway Work Zones*. Report FHWA/RD-80/063. FHWA, U.S. Department of Transportation, Dec. 1980.
2. Z. A. Nemeth and A. K. Rathi. Freeway Work Zone Accident Characteristics. *Transportation Quarterly*, Jan. 1983, pp. 145–159.
3. W. A. Stock and J. J. Wang. Stimulus-Response Lane-Changing Model at Freeway Lane Drops. In *Transportation Research Record 682*, TRB, National Research Council, Washington, D.C., 1978, pp. 64–66.
4. T. Rockwell and Z. Nemeth. *Development of a Driver-Based Method for Evaluating Traffic Control Systems at Construction and Maintenance Zones*. Final Report, EES 581. Ohio State University, Columbus, Oct. 1981.
5. H. W. McGee, W. Moore, and J. H. Sanders. *Decision Sight Distance for Highway Design and Traffic Control Requirements*. Report FHWA-RD-78-78. FHWA, U.S. Department of Transportation, 1978.
6. *Manual on Uniform Traffic Control Devices for Street and Highways*. FHWA, U.S. Department of Transportation, 1978.
7. E. C. Russell. *Building Simulation Models with SIMSCRIPT II.5*. CACI, Los Angeles, Calif., 1981.
8. *Transportation Research Circular 212: Interim Materials on Highway Capacity*. TRB, National Research Council, Washington, D.C., Jan. 1980.
9. F. Haight. *Mathematical Theories of Traffic Flow*. Academic Press, New York, 1963.
10. D. A. Wicks and E. B. Lieberman. *Development and Testing of INTRAS, A Microscopic Freeway Simulation Model*, Vol. I: Program Design, Parameter Calibration and Freeway Dynamics Component Development. Report /RD-80/106. FHWA, U.S. Department of Transportation, 1980.
11. American Public Works Association. *Traffic Control for Construction and Maintenance Work Sites*. Final Report. FHWA, U.S. Department of Transportation, June 1977.
12. A. K. Rathi. *Development of a Microscopic Simulaton Model of Freeway Lane Closures*. Ph.D. dissertation. The Ohio State University, Columbus, 1983.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Abridgment

Integrated Modeling of Freeway Flow and Application to Microcomputers

PANOS G. MICHALOPOULOS AND JAWKUAN LIN

An interactive, menu-driven macroscopic freeway simulation program with graphic capabilities is summarized in this paper. In addition to the employment of personal computers, the program has some attractive features that allow simulation at various levels of complexity. Improved macroscopic modeling specifically developed for the program is used to describe complex phenomena, such as lane changing, merging, diverging, and weaving. The freeway is simulated in an integrated fashion; this implies that the coupling effects of ramps are considered in determining actual entering and exiting flows as well as in following the simultaneous development of queues and propagation of congestion on both the freeway and its ramps. Input to the program is entered interactively and includes conventional traffic parameters, freeway and ramp characteristics (e.g., capacity, free flow speed, jam density), demands (including percentage of exiting volumes at off ramps), and geometric information. Output includes estimation of delays, stops, energy consumption, pollution levels, and other important measures of effectiveness. In addition, two- and three-dimensional plots of speed flow and density are produced for dynamic description of these basic variables in time and space; additional graphics include visual review of the freeway operation during the simulation as well as description of the geometrics, demand patterns, and other input information.

Detailed and realistic analysis of freeway flow dynamics, even in simple situations, can rarely be made analytically or with other convenient tools such as nomographs or design curves. Such analysis is needed in comparing alternative geometric configurations, estimating the effects of improvements, determining the adequacy of traffic management schemes, assessing the impact of control strategies, studying the formation and dissipation of congestion on the freeway and its ramps, and so forth. Although analytical and empirical approximations for answering problems such as these could be made, at best only crude estimates can be expected. More accurate results are usually obtained through simulation. However, this option implies accessibility to software and large computers that are not always readily available to practicing engineers. Furthermore, complexities of software, hardware, or both, make employment of simulation unattractive to the average user.

To make employment of simulation appealing and easily accessible, an interactive, menu-driven microcomputer-based freeway simulation program called KRONOS was recently developed *(1)*. In addition to the employment of personal computers, the attractiveness of the program lies in three areas: (a) its simplicity both in terms of entering the input and interpreting the output, (b) its flexibility in terms of desired accuracy and selecting modeling complexity, and (c) its completeness in terms of graphic capabilities and estimation of the measures of effectiveness. Despite these advantages, the program to this point was available only in a prototype form, that is, it was largely untested and had limitations that are typical of similar experimental software. Since the inception of its initial version, substantial modeling and programming modifications and enhancements have been made. Such improvements include modeling of the freeway and its ramps in an integrated fashion, increased program capabilities, estimation of additional measures of effectiveness, and reduction of memory requirements and execution times. Most importantly, testing against real and simulated data was performed and adjustments were made accordingly.

A major advantage of the program is that it takes into account important traffic phenomena not previously considered by earlier macroscopic programs. Such phenomena include lane changing, merging, diverging, weaving, spillbalk, and friction effects. This was made possible through extensive macroscopic model development and experimentation. In this modeling, the coupling effects of ramps are considered in determining actual entering and exiting flows as well as in following the simultaneous development of queues and propagation of congestion on both the freeway and its ramps.

In the following sections, a brief discussion of the program's modeling and analysis methodology is presented along with its capabilities; more details can be found elsewhere *(2–4)*.

MODELING METHODOLOGY

As mentioned, a number of modeling adjustments were made after testing the initial version of the program with actual and simulated data. Space limitations do not allow detailed model description; this is presented elsewhere *(1–4)*. Suffice it to say that KRONOS IV employs simple continuum models that assume an equilibrium speed-density (or flow-density) relationship. Employment of high-order continuum models, initially allowed in KRONOS I, did not prove to be more effective at least in moderate to heavy flows and were therefore eliminated in subsequent versions. Model implementation is made by discretizing the time space domain in short increments, t and x, respectively, such that $\Delta x/\Delta t > u_f$, where u_f is the free flow speed. Space discretization of a section that includes the most typical freeway components is shown in Figure 1. It should be stressed that this discretization is only mathematical (i.e., it is only done for computational purposes) and not physical.

Although distinct boundaries exist between the freeway components, flow at these boundaries is determined during the

P. G. Michalopoulos and J. Lin, Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis, Minn. 55455.

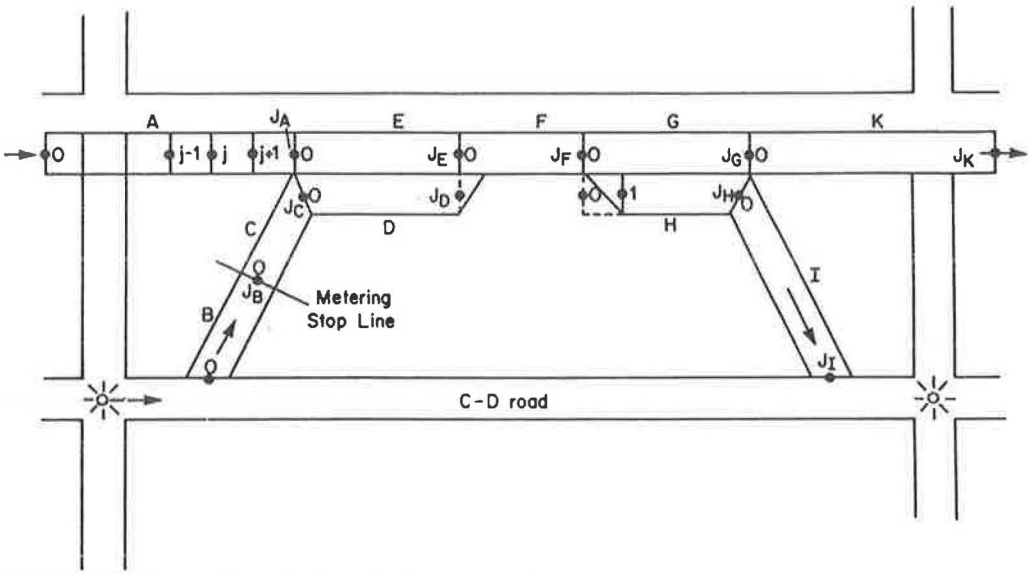


FIGURE 1 **Space discretization of a freeway section.**

solution of the state equations. The only exception is at external boundaries (such as the upstream and downstream ends of the freeway, and the junctions of ramps with the adjacent surface streets) or certain internal ones (such as the metering stop lines and the beginning or ending nodes of deceleration and acceleration lanes, respectively). Flow at these boundaries is specified from the arrival and departure patterns, the metering rates, or other physical considerations.

PROGRAM SUMMARY

The KRONOS IV program employs the previously summarized modeling for dynamically calculating k, u, and Q on each node and segment x. From these basic flow parameters, total travel *(TT)*, total travel time *(TTT)* delay, the duration and extent of congestion (including queue length and size), and energy consumption and pollution levels are derived. These results are summarized by zone (defined as a section between ramps or a merging diverging or weaving area) and by ramp. In ramps the number of vehicles entering and leaving as well as the average and maximum queue length and size are also presented. Both intermediate and final results of the measures of effectiveness are produced. Two and three dimensional plots of k, u, and Q are also produced in order to show the dynamic changes of these variables in space and therefore visualizing the propagation and dissipation of shock waves and congestion on both the freeway and its ramps. For a quicker and easier presentation of the propagation of disturbances along the highway as well as for an easy review of its operation (perhaps after an improvement), the discretized form of the freeway is presented on the graphics screen and each segment is repainted continuously (from blue to red) according to its density as the simulation proceeds.

Additional graphic capabilities of the program are related to the interactive preparation of the input in a fashion requiring minimal reference to the manual. For instance, the geometrics are entered by segmentation and selection of the configuration of each segment from a number of alternatives presented on the graphics screen. Following definition of the geometrics for each segment, the entire freeway is shown for verification.

The remaining input requirements are entered interactively through a series of questions and options as currently done in most expert systems. These inputs are related to the freeway and ramp demand patterns, traffic composition, exiting volumes upstream of entrance ramps, and departure patterns at exit points (off ramps and the downstream end of the freeway). Arrival and departure patterns are also plotted for verification, and can be as complex as desired. The program allows employment of user-specified speed flow models, which are also entered interactively; a default model can be used as an alternative. The u–k model determines capacity of each freeway zone; ramp merging capacity is not needed because the actual number of automobiles entering the freeway is determined during the simulation. Changes to input already entered can be made at any stage during the data entry, and future modifications of a coded facility can easily be accomplished by entering the desired changes only. Finally, input to the program can be short or extended, depending on the accuracy desired. In experimental runs, the time required to enter the necessary data for simulating a three-lane section that is 2 mi long with two ramps for 1 hr using 5-min time varying demands ranged from 10 to 25 min, depending on the user's familiarity with the program.

CAPABILITIES, HARDWARE, AND SOFTWARE REQUIREMENTS

Currently the program can simulate a freeway section up to six lanes wide and approximately 10 miles long. The section can contain up to 20 entrance and 20 exit ramps, while auxiliary lanes between ramps are allowed. A complete list of the program's capabilities includes the following:

1. Reported after each time slice, $N\Delta t$:
   a. Total travel,
   b. Total travel time,

c. Delay,
d. Total arrivals and departures,
e. Queue size and length on each ramp,
f. Energy consumption, and
g. Pollution levels.

2. Reported after simulation is complete:
   a. Plot of speed, flow, or density as a function of distance, time, or both, either two-dimensionally or three-dimensionally;
   b. Plot of congestion areas as a function of time;
   c. Summary of total arrivals and departures;
   d. Number of vehicles in the system at the end of simulation; and
   e. A summary of the measures of effectiveness listed in 1.

3. Color graphics display of density as a function of distance during program execution.

4. Hard copies of all plots (including geometrics) can be obtained through the printer while the input summary and the results of calculations are routed either to the monitor or the printer.

In its current form, the program requires an IBM personal computer with a minimum of 320 K of memory, keyboard, and two floppy disk drives with a minimum storage capacity of 320 K bytes each. It also requires a monochrome monitor, an adaptor, an 80-column printer with graphic capabilities, and a color monitor and color display adaptor. The only software required is the PC-DOS or MS-DOS operating system version 1.1 or higher. Simulation time can be substantially reduced if an IBM-PC 8087 math co-processor board is added to the system. Because the program is available in compiled form, only the software just mentioned is needed. Recompilation of the source code requires the MicroSoft Pascal compiler version 3.2 and access to a disk drive with at least 400 K bytes of disk storage. It is estimated that the total cost for obtaining the hardware just mentioned including options is less than $4,000.

TESTING AND VALIDATION

The early versions of the program were implemented to a number of exemplary situations that covered the entire range of speed flow and density domain and included both entrance and exit ramps as well as multiple lanes. Subsequently, the results were compared with those obtained from microscopic simulation using INTRAS, a recently developed and calibrated program *(5)*. Microscopic simulation during the initial development stage was justified by the need to allow demands and generation rates to fluctuate sufficiently in a controlled environment at relatively short time intervals. A second reason for microscopic data-base generation was the need to impose tractable initial and boundary conditions in order to allow intuitive inspection of the results. The objective of the initial testing was to determine the most appropriate continuum model (among three simple and high-order alternatives) and best numerical algorithm (among a number of options) for use by subsequent versions of the program. In addition, it was necessary to test and adjust the modeling of multilane, merging diverging, and weaving dynamics.

Although the results of the initial tests were encouraging and allowed adjustment of several model parameters, testing with real data was also performed. Such tests required detailed calculations and measurement of k, u, and Q at short time and space increments as well as estimation of the measures of effectiveness at the end of the test period.

Cost considerations did not allow extensive field data collection and analysis; thus, alternative data sources were sought. After requests were made to several sources, the FHWA, U.S. Department of Transportation, provided the most detailed data that could be found at the time of the testing. These data were collected in a recent project at a number of locations around the United States and are microscopic in nature, that is, they include the time individual vehicles cross speed traps placed at strategic locations along selected sections of the freeway *(6)*. Because of the study limitations, the number of test locations was restricted to four, representing typical freeway components, namely, pipeline, merging, diverging, and weaving sections *(4)*.

Comparisons with the field data led to similar conclusions with the simulated data base. Interestingly, it was found that discontinuous equilibrium u–k models generally increased accuracy, and this was more pronounced in the diverging case. However, because derivation of such relationship is rather tedious for most practical applications, a generalized one was derived from the available data in all sites. To demonstrate that the program can realistically handle larger freeways, a real freeway section approximately 10 miles long, containing 19 ramps and 3 basic lanes, was simulated for 90 min; the results favorably compared with earlier data *(4)*.

CLOSING REMARKS

The most interesting feature of the KRONOS IV program is its ability to run on a microcomputer while its results are in agreement with both real and simulated data. This was made possible by the simplicity of the modeling and numerical methodologies developed, which allow quick and accurate calculation of the basic flow variables in both time and space. Incidentally, it is worth mentioning that three-dimensional models that explicitly take street width into account (along with time and length) were also developed *(2)*, but were later excluded for reducing the computational effort. Based on discussions and comments received to this point, the authors feel compelled to repeat that the proposed discretization is only mathematical, not physical; this implies that both data and field measurements should only be made in larger sections (zones), as in most practical applications. Of equal importance is the integrated treatment of ramps, acceleration and deceleration lanes, merging and diverging, and weaving areas, as well as the inclusion of lane-changing effects. This is a feature that along with the interactive graphics has been missing from earlier macroscopic programs.

Despite the program's advantages, which should encourage employment of simulation by practicing engineers, certain shortcomings should be recognized. The most important one is related to execution time, which can be very long as the size and complexity of the freeway increases. For instance, execution time for a three-lane, two-ramp section for 1 hr, including

detailed lane-by-lane analysis and all graphic options, is approximately 18 to 19 hr, which can be considered long compared with the much faster performance of large computers.

However, recent hardware advances (such as the introduction of the IBM-PC/AT) as well as new software development (such as the new Turbo-Pascal) should reduce execution time considerably. Additional enhancements that should make the program more attractive are possible and some are planned. Such improvements are related to inclusion of left-hand-side ramps, collector distributor roads, incidents, high-occupancy-vehicle–priority treatment, demand diversion, generation of origin-destination patterns after implementation of a new management policy, estimation of optimal ramp metering lanes, and so forth. Such improvements will allow treatment of most freeway operational problems one is likely to encounter in practice.

ACKNOWLEDGMENTS

Financial support for completing this project was provided by the National Science Foundation, IBM, and the Minnesota Department of Transportation.

REFERENCES

1. P. G. Michalopoulos. A Dynamic Freeway Simulation Program for Personal Computers. In *Transportation Research Record 971*, TRB, National Research Council, Washington D. C., 1984, pp. 68–79.
2. P. G. Michalopoulos, D. E. Beskos, and Y. Yamauchi. Multilane Traffic Flow Dynamics: Some Macroscopic Considerations. *Transportation Research*, Vol. 18B, No. 4/5, 1984, pp. 377–395.
3. P. G. Michalopoulos. Integrated Modelling of Freeway Flow and Application to Microcomputers. *Traffic Engineering and Control*, 1986 (in press).
4. P. G. Michalopoulos, R. Plum, and J. Lin. *KRONOS IV: An Interactive Freeway Simulation Program for Personal Computers*. Final Report. Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis, 1985.
5. D. A. Wicks and E. B. Lieberman. *Development and Testing of INTRAS, A Microscopic Freeway Simulation Model*, Vols. 1-4. Report DOT-FH-11-8502, FHWA, U. S. Department of Transportation, 1980.
6. E. L. Seguin. *Urban Freeway Truck Characteristics: Data Base Documentation*. Interim Report. FHWA Project DT-FH-61-80-C-00106. FHWA, U. S. Department of Transportation, 1982.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Statistical Analysis of Output Ratios in Traffic Simulation

A. V. GAFARIAN AND A. HALATI

Simulation models are increasingly becoming the most convenient tool for traffic studies. Users of such models need valid statistical methods to draw correct inferences. Presented in this paper is one such method applicable to several important traffic parameters. The motivation for this research arose from a study sponsored by the FHWA, U. S. Department of Transportation, to develop statistical guidelines for simulation experiments with traffic models. NETSIM, widely used for simulating vehicular traffic flow on urban streets, was used in the study. The output of the NETSIM model includes estimates of average speed, average delay per vehicle, and average travel time per vehicle mile. Because NETSIM uses the ratio of sample means to estimate these parameters, a situation exists that involves the ratios of observations that are in fact autocorrelated and cross correlated. In this paper, the efficacy of the ratio of sample means (used in NETSIM) as an estimator of the ratio of steady state means is discussed. Monte Carlo experiments have demonstrated that the user of the NETSIM model, in estimating these parameters from the model output, must apply statistical techniques based on ratio estimators. A technique that provides a measure of the accuracy of the estimate with a confidence interval is developed and demonstrated. The efficacy of the method is assessed through Monte Carlo experiments. The method is easy to use and can be applied just as readily to field data. It can be extended to the comparison of model outputs to field observations for simulation validation studies.

NETSIM is a widely accepted simulation tool for simulating traffic behavior on urban networks *(1, 2)*. The basic input requirements of the model are the network geometry, signalization information, and traffic counts, which consist of both input flow rates and turning movements. The standard output of the NETSIM model includes estimates of important traffic parameters such as

- Total vehicle minutes of travel time,
- Number of vehicles discharged,
- Total vehicle miles of travel distance,
- Average travel time per vehicle
- Average travel time per vehicle mile,
- Average speed, and
- Average delay time per vehicle.

The estimates of the traffic parameters are provided both on a link-by-link basis (links represent a one-way direction of flow on a street typically between two successive stop bars) and on a network basis.

A. V. Gafarian, Department of Industrial and Systems Engineering, University of Southern California, University Park, Los Angeles, Calif. 90089. A. Halati, California State Polytechnic University at Pomona, Calif. 91768.

It will be demonstrated that (a) each of the last four measures of effectiveness (MOEs) is a parameter that is the ratio of means of two random variables X and Y (X and Y are used generically here), that is, the MOE itself is μ_X/μ_Y, and (b) the natural estimate NETSIM provides is the ratio of the sample means of the X and Y random variables. This will be done in some detail for two parameters: average speed and average delay time per vehicle. The extension to two other parameters, average travel time per vehicle and average travel time per vehicle mile, will be obvious.

The discussion begins by noting that NETSIM is a stochastic microscopic traffic simulation model with a basic sampling interval of 1 second. Thus, the status of each individual vehicle is sampled at the rate of once every second and all required statistics are updated at the end of every second.

Example 1: Average Speed on a Link. In the simplest case, after the initial warm-up period, NETSIM produces for each 1-sec Δt time period the following observations on the totality of vehicles exiting the link during the time period under consideration: (a) number of vehicle miles in the link (equal to link length times the number exiting) and (b) vehicle minutes in the link (amount of time spent by all vehicles traversing the link). Running the model after warm-up for some integral multiple T of $\Delta t = 1$ provides the following as an estimate of average speed:

$$\sum_{j=1}^{T} (\text{vehicle miles in } j\text{th } \Delta t \text{ time period}) \div \sum_{j=1}^{T} (\text{vehicle minutes in } j\text{th } \Delta t \text{ time period})$$

To understand why there is a ratio of two means $\bar{X}$ and $\bar{Y}$ that are estimates of μ_X and μ_Y, more work needs to be done. First, observe that as things stand now these observations, in both the numerator and the denominator, are not identically distributed. For example, because travel distance on a link in NETSIM is proportional to the number of vehicles discharged, during the red interval of the downstream signal the travel distance on the link will be accumulated at a low rate; during the early portion of the green interval, the travel distance will be accumulated at a large rate while the queue is dissipating.

Thus, there are observations on random variables that do not even have the same mean let alone the same distribution. Therefore, dividing the numerator and the denominator in the above expression by T does not give an estimate of the mean of any well-defined random variable.

It should be noted at the outset that it is important to deal with identically distributed observations because the problem

of making valid statistical statements becomes tractable. To achieve identically distributed observations in both the numerator and the denominator of the ratio just given (so that dividing the numerator and the denominator by the number of observations gives estimates of the numerator and denominator means) is easy. All the observations for each $\Delta t = 1$ during one cycle of the link's downstream signal are summed. Thus, if the cycle length is 60 sec, then 60 sets of vehicle miles are added to produce one observation of vehicle miles. Likewise, the same is done with vehicle minutes in the denominator. A little reflection shows that these sums, from cycle to cycle, are certainly identically distributed after the warm-up. The remaining MOEs can be treated similarly.

Hereafter, it will be assumed that the collection interval will equal the downstream cycle length of each link. If there are two cycle lengths present in the network, for example, 60 and 90 sec, then running the model for (180)k seconds would provide 3k cycles' worth of observations for 60-sec links and 2k cycles' worth of observations for 90-sec cycle links.

To continue, take the above ratio, group the data as described, and end up with the following ratio for the estimate of average speed:

$$(X_1 + X_2 + \ldots + X_n)/(Y_1 + Y_2 + \ldots + Y_n)$$

where

X_i = accrued vehicle miles of vehicles departing during ith cycle, $i = 1,2, \ldots, n$; and
Y_i = accrued vehicle minutes in the link of vehicles departing during ith cycle, $i = 1,2, \ldots, n$.

Because the X_i's and Y_i's are identically distributed, this may be written

$$[(X_1 + X_2 + \ldots + X_n)/n]/[(Y_1 + Y_2 + \ldots + Y_n)/n] = \bar{X} / \bar{Y}$$

As $n \to \infty$, the numerator and denominator converge to μ_X and μ_Y, respectively, both with probability 1, where μ_X is the average vehicle miles per cycle and μ_Y is the average vehicle minutes per cycle. Thus, the problem of estimating link average speed is the same as estimating μ_X/μ_Y, the ratio of two means.

Example 2: Average Delay on a Link. Here again the following would be an estimate of average delay:

$\sum_{j=1}^{T}$ (accrued delay of vehicles departing the link during jth Δt time period) +
$\sum_{j=1}^{T}$ (number of vehicles departing the link during jth Δt time period)

As in Example 1, the same arguments could be used to get X_i's and Y_i's each identically distributed where

X_i = accrued delay of vehicles departing during the ith cycle, $i = 1,2, \ldots, n$; and
Y_i = N_i = number of vehicles departing during the ith cycle, $i = 1,2, \ldots, n$, (note that Y_i in this case is an integer-valued random variable).

and to produce an estimate of average delay $\sum_{i=1}^{n} X_i / \sum_{i=1}^{n} N_i = \bar{X} / \bar{N}$, which converges with probability 1 to μ_X/μ_N as $n \to \infty$. So again the ratio of two means is estimated. In this case, μ_X is the average delay per cycle and μ_N is the average number of vehicles discharged per cycle.

The principal objective of this paper is to develop a statistically valid method for using $\bar{X} / \bar{Y}$ as a point estimate for μ_X/μ_Y and to provide, with a confidence interval, a measure of its accuracy. What X and Y are depends on the particular MOE being estimated.

STATISTICAL PROPERTIES OF THE OBSERVATIONS

In this section, some important statistical properties of the observations will be described.

Observations Tend To Be Normal

This property follows from the fact they are sums of random variables, obtained by adding up all the individual observations for each Δt. Thus, the Central Limit Theorem, which holds for fairly unrestrictive conditions (even when the variables being added are not identically distributed or independent), comes into play and it can be stated that asymptotic normality is obtained. This includes integer-valued observations, such as the number of vehicles discharged during a cycle length.

In this connection, it should be mentioned that the method developed in the paper is based on the t-statistic and that this statistic is robust with respect to normality; that is, inferences using it are not seriously invalidated by the violation of the normality assumption. This will be demonstrated in the Monte Carlo experiment presented later in the paper.

Observations Are Not Independent

The observations of travel time and travel distance, for example, are each autocorrelated. Figure 1 shows estimates of autocorrelation for travel time on a link of a simple star network consisting of essentially an isolated four-legged intersection with pretimed signal control. The simulation run consisted of 130 cycles of a common signal cycle length of 80 sec (i.e., 10,400 sec). Estimates of autocorrelation

$$r_{XX}(k) = c_{XX}(k)/c_{XX}(0) \quad k = 0, 1, \ldots, n/10$$

were obtained by using

$$c_{XX}(k) = (1/n)\sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X}) \quad k = 0, 1, \ldots, n/10$$

where

$r_{XX}(k)$ = sample autocorrelation of the X series for lag k,
$c_{XX}(k)$ = sample autocovariance of the X series for lag k,
n = number of cycles that made up the simulation run, in this case 130, and

X_i = observed value of total travel time during cycle i.

The maximum lag was restricted to $n/10$ to obtain accurate estimates of the autocorrelations.

If it is assumed that travel time observations are independently, identically, and normally distributed random variables, the standard deviation of the autocorrelation estimates are approximately equal to $\sqrt{1/n}$, (*3*, pp.34, 35). In this case, the standard deviation would be approximately .0877. Because the estimate of the first lag autocorrelation is .321, almost four times the standard deviation, it can be concluded that the first lag correlation is not zero. Moreover, there is strong indication that there is autocorrelation up to lag 10. Thus, it is reasonable to assume that successive travel times are autocorrelated.

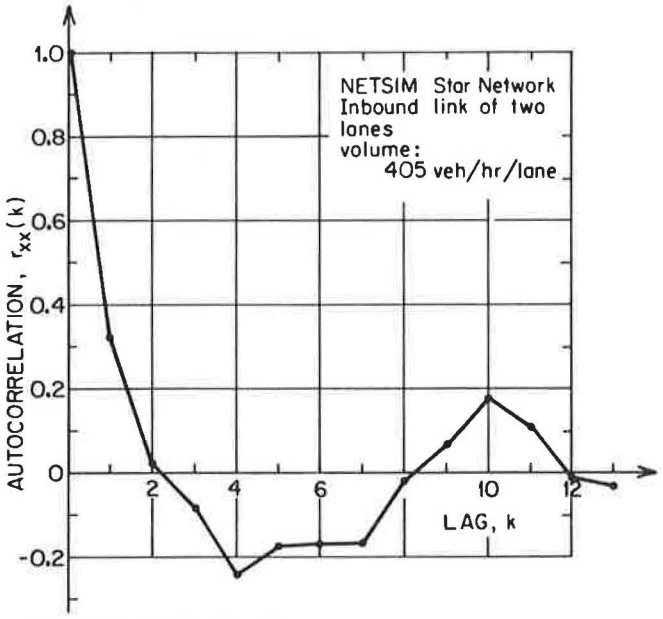


FIGURE 1 Estimates of autocorrelation for travel time on a link of a simple star network.

Observations of Random Variables of the Numerator and Denominator Are Cross Correlated

Figure 2, for example, shows the cross-correlation estimates between the total travel time per cycle and the total travel distance per cycle. Estimates of cross correlation

$$r_{XY}(k) = c_{XY}(k)/c_{XX}(0)c_{YY}(0) \quad k = 0, \pm 1, \ldots, \pm(n/10)$$

were obtained using

$$c_{XY}(k) = (1/n) \sum_{i=1}^{n-k} (X_i - \bar{X})(Y_{i+k} - \bar{Y}) \quad k = 0, \quad 1, \ldots, n/10$$

$$= (1/n) \sum_{i=1}^{n+k} (X_{i-k} - \bar{X})(Y_i - \bar{Y}) \quad k = -1, -2, \ldots, n/10$$

where

$r_{XY}(k)$ = sample cross correlation of the X and Y series for lag k,

$c_{XY}(k)$ = sample cross covariance of the X (travel time) and Y (travel distance) series for lag k,

n = number of cycles that made up the simulation run, in this case 130,

X_i = travel time during cycle i, and

Y_i = travel distance during cycle i.

Again the maximum lag was restricted to $\pm n/10$ to obtain accurate estimates of the cross correlations.

Note that the cross covariance has both positive lags (where Y leads X) and negative lags (where X leads Y); and that, in general, $c_{XY}(k) \neq c_{XY}(-k)$. This is not the case for the autocovariance, where $c_{XX}(k) = c_{XX}(-k)$ [or $c_{YY}(k) = c_{YY}(-k)$]. The large cross-correlation estimate of lag 0 is expected because a large observation for travel distance indicates that a large number of vehicles have traversed the link and thus a large value of travel time has been incurred. In addition, significant cross correlation at larger lags is also observed.

PROBLEM DEFINITION

There are two common methods for performing simulation experiments, and the problem will be defined for each of these methods.

Method 1: A Single Long Run

The first method consists of running the simulation model for a long duration and using the observations generated in this single, continuous, long run to estimate the parameters of interest and to obtain a measure of the accuracy of the estimate.

In the case of the NETSIM model, as it pertains to the traffic parameters that are estimated as the ratio of two random variables (which in this case happens to be means), it was demonstrated in the preceding section that successive observations obtained on the random variables (at the end of each cycle) are autocorrelated and cross correlated. In the presence of these correlations, estimating the parameters from a single contin-

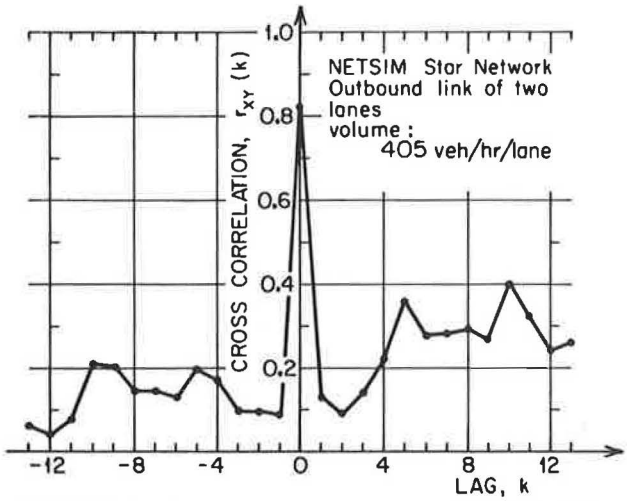


FIGURE 2 Cross-correlation estimates between total travel time per cycle and total travel distance per cycle.

uous run of the NETSIM model may be defined as the following statistical problem.

Let (X,Y) be a bivariate random variable and suppose a sequence of identically distributed observations $[(X_1,Y_1), (X_2,Y_2), \ldots, (X_n,Y_n)]$ has been obtained. The X_i's and Y_i's correspond to the numerator and denominator observations, respectively. Furthermore, assume the following:

$E[X] = \mu_X \quad \text{Var}[X] = \sigma_X^2$

$E[Y] = \mu_Y \quad \text{Var}[Y] = \sigma_Y^2$

$E[(X_i - \mu_X)(X_{i+k} - \mu_X)]/\sigma_X^2 \neq 0 \quad$ for all k

(observations are not independent);

$E[(Y_i - \mu_Y)(Y_{i+k} - \mu_Y)]/\sigma_Y^2 \neq 0 \quad$ for all k

(observations are not independent);

$E[(X_i - \mu_X)(Y_{i+k} - \mu_Y)]/\sigma_X\sigma_Y \neq 0 \quad k \geq 0$

(observations of random variables of the numerator and denominator are cross correlated);

$E[(X_{i-k} - \mu_X)(Y_i - \mu_Y)]/\sigma_X\sigma_Y \neq 0 \quad k < 0$

(observations of random variables of the numerator and denominator are cross correlated).

The problem then becomes that of using this information to estimate $R = \mu_X/\mu_Y$ and of assessing the accuracy of the estimate by constructing a confidence interval.

The two examples of this generic problem discussed in the introduction were average link speed and mean delay:

1. Average link speed. Here the point estimate is $\bar{X} / \bar{Y}$, where X_i is accrued vehicle miles of vehicles departing during the ith cycle and Y_i is accrued vehicle minutes in the link of departing vehicles during the ith cycle, $i = 1,2, \ldots, n$.
2. Mean delay. Here the point estimate is $\bar{X} / \bar{N}$, where X_i is accrued delay of vehicles departing the link during the ith cycle and N_i is the number of vehicles departing during the ith cycle, $i = 1,2, \ldots, n$.

The problem of using the observations from a single run and developing a confidence interval for μ_X/μ_Y (average speed in the first example) or μ_X/μ_N (mean delay in the second example) is extremely complex and involves estimating autocorrelations of the two numerator and denominator variables and the cross correlations of the numerator variables with the denominator variables [see Halati (*4*, pp.65–69)]. This requires an extremely long run to get reliable estimates of all the needed correlations, as well as to reduce the inherent bias present in the estimate $\bar{X} / \bar{Y}$ [Halati (*4*, p.63)].

In addition, the use of the method is predicated on collecting observations that are identically distributed. It was noted that identically distributed observations may be obtained by summing the statistics, collected at the end of each 1-sec sampling interval, over the period of one cycle length. This is obviously applicable if the network consists solely of pretimed controls. When actuated controls are present, there is no immediate alternative for producing identically distributed observations.

The method of independent replications, which will be discussed next, does not have this disadvantage. Each replication will result in a single observation that is the sum of the statistics over the duration of each run. The notion of the cycle will not be needed.

Method 2: Several Independent Replications

The second method for conducting simulation experiments is to perform independent replications. In this method, repeated runs of the model are performed in such a way that the output of the model in each run is independent of the others by using a different random generator seed in each run.

In this method, the great difficulty of getting reliable estimates of all the autocorrelations and cross correlation among successive observations is circumvented. However, the method has the disadvantage of requiring a warm-up time for each replication during which no data may be collected.

In the case of the NETSIM model, and again as it pertains to the analysis of those traffic parameters that are the ratio of two random variables, the problem of estimating these parameters and assessing the accuracy of the estimate may now be defined statistically in the following way.

Suppose n independent replications of the NETSIM model are performed and also assume that each run has a prescribed duration of k cycles. In this case the observations, (X_i, Y_i), $i = 1,2, \ldots, n$, would be the cumulative values of the observed statistics at the end of each run and

- X_i, $i = 1,2, \ldots, n$, would be a sequence of IID observations,
- Y_i, $i = 1,2, \ldots, n$, would be a sequence of IID observations,
- $E[X_i] = k\mu_X$ and $E[Y_i] = k\mu_Y$ because now X_i and Y_i are the cumulative values obtained by adding the statistics over k cycles, and
- The only cross correlation present is between X_i and Y_i.

The problem in this form is that of using the cumulative statistics $\{(X_i,Y_i), i = 1,2, \ldots, n\}$ to estimate $R = \mu_X/\mu_Y$ and assess the accuracy of the estimate.

Two points should be noted here. The first point is that $\sum_{i=1}^n X_i/\sum_{i=1}^n Y_i$ is still an estimator of R even though the X_i's and Y_i's are cumulative values when independent replications are performed. This is because, with probability 1,

$$\lim_{n\to\infty} \left(\sum_{i=1}^{n} X_i / \sum_{i=1}^{n} Y_i \right) = k\mu_X/k\mu_Y = R$$

The second point is that the normality assumption discussed in the section on statistical properties of the observations becomes better. This is because each X_i and Y_i is now the sum of a larger number of observations.

Thus, in the first example above

X_i = accrued travel distance of vehicles departing the link during the ith replication, $i = 1,2, \ldots, n$, and

Y_i = accrued vehicle minutes in the link of vehicles departing the link during the ith replication, $i = 1,2, \ldots, n$

and for an estimate of average speed for the link one would take

$$\sum_{i=1}^{n} X_i / \sum_{i=1}^{n} Y_i = \bar{X} / \bar{Y}$$

In the second example,

X_i = total accrued delay of the ith replication, $i = 1,2, \ldots, n$, and

Y_i = N_i = total number of vehicles departing the link during the ith replication, $i = 1,2, \ldots, n$, and for an estimate of mean delay one would take

$$\sum_{i=1}^{n} X_i / \sum_{i=1}^{n} N_i = \bar{X} / \bar{N}$$

To reiterate the important points with respect to independent replications, there are no autocorrelations among the X_i's or the Y_i's and there is only a cross correlation between X_i and Y_i. It is the problem in this form that will be studied in this paper.

PERFORMANCE DEGRADATION WHEN OBSERVATIONS FROM INDEPENDENT REPLICATIONS ARE NOT TREATED AS A RATIO OF RANDOM VARIABLES

Before proceeding with the development of the procedure for computing a confidence interval for μ_X/μ_Y by using the point estimator $\sum_{i=1}^{n} X_i / \sum_{i=1}^{n} Y_i = \bar{X} / \bar{Y}$, consider an important question of the degradation that occurs when the problem is not treated as one in ratio estimation. The reason for doing this is that one might easily be tempted to develop a confidence interval for, say, mean speed from independent replications by a method that goes as follows. Because each replication gives an independent estimate of mean speed $Z_i = X_i/Y_i$, $i = 1,2, \ldots, n$, one would estimate mean speed as

$$\bar{Z} = (1/n) \sum_{i=1}^{n} Z_i$$

and assess the accuracy of this estimate with a $(1 - \alpha) \times 100$ percent classical confidence interval

$$\left\{ \left[\bar{Z} - t_{1-(\alpha/2),(n-1)} \right] S/(n)^{1/2}, \left[\bar{Z} + t_{1-(\alpha/2),(n-1)} \right] S/(n)^{1/2} \right\}$$

where $t_{1-(\alpha/2),(n-1)}$ is the upper $\alpha/2$ point of the t-statistic with $n - 1$ degrees of freedom and

$$S = [1/(n-1)] \sum_{i=1}^{n} (Z_i - \bar{Z})^2$$

In effect, the observations from the model and field were treated as just described in the validation studies performed on NETSIM (2, pp.147–248). The thing that is wrong with this procedure is that one is really estimating (or doing hypothesis testing) on $E[X/Y]$ and not μ_X/μ_Y. It is well known that in general $E[X/Y] \neq \mu_X/\mu_Y$. To demonstrate that this approach to the analysis of ratio estimators of the NETSIM model may produce results that are greatly in error, a Monte Carlo study was conducted. The study consisted of generating bivariate normal random variables (X_i,Y_i) for selected sample sizes n with the variance-covariance matrix

$$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

X_i and Y_i correspond to the numerator and the denominator observations respectively of the ith replication and n corresponds to the number of replications. μ_X was chosen as 100 and μ_Y as 5. Thus, the known and true value of the ratio of means was 20.

For each sample size n and using the above procedure, a 95 percent confidence interval was constructed. To assess the goodness of the confidence interval, the experiment was repeated 500 times for each sample size and the following four measures of effectiveness on the behavior of the constructed confidence intervals were computed:

1. Coverage probability. This measure of effectiveness is the fraction of the confidence interval produced in the 500 repetitions of the experiment that covered the true value of the ratio of the means, which was 20. Closeness of this value to .95 is obviously a desired property of the method.
2. Coefficient of variation of coverage probability. This statistic is the ratio of the standard deviation of the estimate of coverage probability to the estimated coverage probability. It is a measure of how good the estimate of coverage probability is—the smaller the value the better the estimate. Thus, .010 for $n = 5$ means that the standard deviation of the estimate is only 1 percent of the estimate.
3. Average confidence interval length. In each repetition of the experiment, the length of the constructed confidence interval was recorded. This statistic represents the average of the recorded confidence interval lengths over the 500 repetitions. Obviously, the smaller the length the better.
4. Coefficient of variation of expected confidence interval length. This is the ratio of the standard deviation of the estimate of the average confidence interval length to the average confidence interval length. Again, it is a measure of how good the estimate of average confidence interval length is. For $n = 5$, .023 means that the standard deviation of the estimate is about 2.3 percent of the estimate.

The study was conducted for replication sizes of 5,6,7,8,9,10,20,50,100, and 200 observations per replication. The results are given in Table 1.

To obtain a basis for comparison of the results, the study was repeated identically by using the proposed method (to be developed in the next section) of analysis. The results of that study are given in Table 2.

It is noted that the coverage probabilities are substantially reduced by using the incorrect method of analysis described at the beginning of this section. It is seen here that as the number

TABLE 1 RESULTS OBTAINED FROM MONTE CARLO EXPERIMENTS USING INCORRECT METHOD

No. of Replications	Coverage Probability	Coefficient of Variation of Coverage Probability	Avg Confidence Interval Length	Coefficient of Variation of Avg Confidence Interval Length
5	.950	.010	10.48	.023
6	.936	.009	9.04	.022
7	.938	.011	7.80	.021
8	.946	.010	7.48	.022
9	.934	.011	6.82	.020
10	.936	.011	6.48	.020
20	.904	.013	4.34	.014
50	.754	.019	2.70	.010
100	.574	.022	1.92	.009
200	.206	.018	1.32	.006

of replications increases smaller confidence intervals result, and they begin to miss μ_X/μ_Y in increasing numbers {because they are really covering $E[X/Y] \neq (\mu_X/\mu_Y)$}. However, the proposed method of analysis, which is based on a ratio estimation technique, produced coverage probabilities close to the desired 95 percent percent for all sample sizes. At smaller replication sizes, when both methods appear to have coverage probabilities close to the desired .95, the proposed method consistently resulted in more precise (lower coefficient of variation) and smaller average confidence interval lengths.

PROPOSED METHOD

How a confidence interval may be developed for μ_X/μ_Y based on observations $[(X_i, Y_i), i = 1,2, \ldots, n]$ obtained from each independent replication of the model will now be considered. Keep in mind as an example that X_i is total travel distance in vehicle-miles and Y_i is total travel time in vehicle-minutes (or X_i could be total delay and $Y_i = N_i$ total number of vehicles discharged). The method is a small sample extension of the Fieller method *(5)*.

Suppose a confidence interval is wanted for $R = \mu_X/\mu_Y$. Then the estimator would be

$$\hat{R} = \sum_{i=1}^{n} X_i / \sum_{i=1}^{n} Y_i = (1/n) \sum_{i=1}^{n} X_i / (1/n) \sum_{i=1}^{n} Y_i = \bar{X} / \bar{Y}$$

Next, a new variable is considered:

$$Z_i = X_i - RY_i$$

and then

$$\bar{Z} = \bar{X} - R\,\bar{Y}$$

If it is assumed that X_i and Y_i are normally distributed (this assumption has been discussed in the section on statistical properties of the observations), then Z_i and $\bar{Z}$ will be normally distributed. Because $E[Z_i] = E[\bar{Z}] = 0$,

$$\bar{Z} / \left\{ (1/n) \cdot [1/(n-1)] \sum_{i=1}^{n} (Z_i - \bar{Z})^2 \right\}^{1/2}$$

TABLE 2 RESULTS OBTAINED FROM MONTE CARLO EXPERIMENTS USING PROPOSED METHOD

No. of Replications	Coverage Probability	Coefficient of Variation of Coverage Probability	Avg Confidence Interval Length	Coefficient of Variation of Avg Confidence Interval Length
5	.940	.011	10.48	.023
6	.928	.012	8.54	.019
7	.938	.011	7.18	.016
8	.944	.010	6.68	.016
9	.936	.011	6.06	.014
10	.942	.010	5.70	.013
20	.942	.010	3.70	.009
50	.960	.009	2.24	.005
100	.968	.008	1.56	.004
200	.938	.011	1.08	.003

has a Student t-distribution with $n-1$ degrees of freedom. But

$$[1/(n-1)] \sum_{i=1}^{n} (Z_i - \bar{Z})^2 = [1/(n-1)] \sum_{i=1}^{n} (X_i - RY_i - \bar{X} + R\bar{Y})^2$$

$$= [1/(n-1)] \sum_{i=1}^{n} [(X_i - \bar{X}) - R(Y_i - \bar{Y})]^2$$

$$= [1/(n-1)] \sum_{i=1}^{n} [(X_i - \bar{X})^2 + R^2(Y_i - \bar{Y})^2 - 2R(X_i - \bar{X})(Y_i - \bar{Y})]$$

$$= (S_X^2 + R^2 S_Y^2 - 2RS_{XY})$$

where

$$S_X^2 = [1/(n-1)] \sum_{i=1}^{n} (X_i - \bar{X})^2 = \text{sample variance of the } X_i\text{'s}$$

$$S_Y^2 = [1/(n-1)] \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \text{sample variance of the } Y_i\text{'s}$$

$$S_{XY} = [1/(n-1)] \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = \text{sample covariance of the } (X_i, Y_i)\text{'s}$$

Hence at the $(1-\alpha)$ level,

$$Pr\left(\left\{(\bar{X} - R\bar{Y}) / \left[(1/n)(S_X^2 + R^2 S_Y^2 - 2RS_{XY})\right]^{1/2}\right\} \le t_{1-(\alpha/2),n-1}\right) = 1 - \alpha$$

Note that when both sides of the argument in the above probability statement are squared, the result is a quadratic inequality in the unknown $R = \mu_X/\mu_Y$, the known estimates $\bar{X}$, $\bar{Y}$, S_X^2, S_Y^2, S_{XY}, and $t_{1-(\alpha/2),(n-1)}$. The roots of this quadratic inequality are

$$\left[\bar{X}\bar{Y} - g(\alpha)S_{XY}\right] \pm \left(\left[\bar{X}\bar{Y} - g(\alpha)S_{XY}\right]^2 - \{[\bar{Y}^2 - g(\alpha)S_Y^2][\bar{X}^2 - g(\alpha)S_X^2]\}\right)^{1/2} \Big] / \left[\bar{Y}^2 - g(\alpha)S_Y^2\right]$$

where

$$g(\alpha) = t_1^2 - \alpha/2;\ n - 1/n$$

The $(1-\alpha) \times 100$ percent confidence interval is then (r_1, r_2), where r_1 is the smaller root and r_2 is the larger root. It should be noted that $\bar{X}/\bar{Y}$ is not the midpoint of the confidence interval.

The efficacy of the proposed method in any application depends on how well the assumptions are met; namely, system in steady state, normality of the numerator and the denominator observations, and independent replications.

Steady state in NETSIM is achieved by a warm-up procedure that appears to work. Independent replications are achieved by starting each run with a different random number generator seed. The only assumption that is approximately met in applications is that of normality, and it is claimed that the method is not sensitive to this requirement.

To demonstrate the method's robustness to this assumption, it was applied to the M/M/1 queuing system to estimate mean delay (excluding service time) per customer. This is a single-server system with exponential interarrival and service times. This system was selected because it is known that delay in queue is extremely nonnormal.

The arrival rate was taken to be $\lambda = 36$ customers per hour and the service rate $\mu = 40$ customers per hour for a traffic intensity $\rho = \lambda/\mu = 36/40 = 0.90$. Each replication was started in steady state. This was accomplished by having an initial number of customers in the system obtained from sampling the steady state probability mass function given by $p_n = (1-\rho)\rho^n$, $n = 0,1,2, \ldots$.

Each replication consisted of 2 hours of simulated time. The numerator and denominator recordings were, respectively, X_i = accumulated delay of departing customers, and $Y_i = N_i$ = number of departing customers, $i = 1,2, \ldots, n$. Thus, the result is a ratio estimation situation in which $\sum_{i=1}^{n} X_i / \sum_{i=1}^{n} N_i$ converges as $n \to \infty$ to μ_X/μ_N, where μ_X is the expected total delay per unit time and μ_N is the expected number of departures per unit time.

The number of replications was selected to be 5,6,7,8,9,10,20, and 40. For each replication size, the experiment was repeated 500 times and for each repetition a 95 percent confidence interval was constructed by the proposed method. Because for this system the true value of the steady state mean delay is known to be $\lambda/[\mu(\mu - \lambda)]$ (in this case 13.5 min), an estimate of coverage probability is the fraction of the 500 repetitions, which cover the true value. Also, estimates of the average confidence interval length and of the coefficients of variation of both coverage probability and average confidence interval length were computed and the results are given in Table 3.

The experiment was then repeated with an arrival rate of $\lambda = 32$ customers per hour and a service rate of $\mu = 40$ customers per hour for a traffic intensity of $\rho = \lambda/\mu = 32/40 = 0.80$. These results are given in Table 4.

Looking at the data in these two tables, some degradation of coverage probability due to the extreme nonnormality of the data can be seen, but the results are not bad. The worst case is an 81.4 percent coverage when 95 percent was expected, for $n = 5$ and $\rho = 0.8$. However, it improves rapidly and for 40 replications it is up 93.6 percent. The situation is considerably better for $\rho = 0.9$. It should be noted that the problem of nonnormality can be ameliorated by making the replications longer.

SUMMARY AND CONCLUSIONS

In this paper a number of important traffic parameters were identified as being ratio of means of two random variables. These parameters may be estimated by the ratio of sample means. In particular, it was noted that NETSIM uses this type of estimator. It was demonstrated that, in general, the numerator and denominator random variables that comprise the sample means are autocorrelated and cross correlated. Therefore,

TABLE 3 M/M/1 QUEUING SYSTEM, AVERAGE DELAY PER CUSTOMER (ρ = 0.9)

No. of Replications	Coverage Probability	Coefficient of Variation of Coverage Probability	Avg Confidence Interval Length	Coefficient of Variation of Avg Confidence Interval Length
5	.882	.016	29.86	.030
6	.876	.017	26.90	.028
7	.860	.018	23.18	.027
8	.888	.016	21.94	.023
9	.864	.018	18.68	.023
10	.874	.017	18.36	.023
20	.920	.013	12.96	.017
40	.950	.010	8.66	.011

TABLE 4 M/M/1 QUEUING SYSTEM, AVERAGE DELAY PER CUSTOMER (ρ = 0.8)

No. of Replications	Coverage Probability	Coefficient of Variation of Coverage Probability	Avg Confidence Interval Length	Coefficient of Variation of Avg Confidence Interval Length
5	.814	.021	12.06	.036
6	.874	.017	10.02	.030
7	.856	.018	9.80	.032
8	.870	.017	7.72	.028
9	.892	.016	8.04	.025
10	.872	.017	7.34	.025
20	.924	.013	5.08	.025
40	.936	.012	2.74	.014

obtaining estimates from a single continuous set of observations and assessing the accuracy of the estimate by a confidence interval is an extremely complicated statistical problem.

Because the method of independent replications simplifies the problem considerably, a method based on it was developed and its efficacy was demonstrated through Monte Carlo experiments. The method may be applied to the estimation of parameters and assessment of the accuracy of the estimates from field data, to the analysis of traffic simulation outputs, and to the comparison of field data with simulated data for validation studies.

REFERENCES

1. E. Lieberman, R. D. Worrall, D. Wicks, and J. Woo. *NETSIM Model*, Vol. 4: User's Guide. Report FHWA-RD-77-44. FHWA, U. S. Department of Transportation, Oct. 1977.
2. *Network Flow Simulation for Urban Traffic Control System.* General Applied Science Laboratories, Inc., Final Report FH-11-7462-2. Peat, Marwick, Mitchell and Co., Washington, D. C., 1972.
3. G. E. P. Box and G. M. Jenkins. *Time Series Analysis—Forecasting and Control.* Holden-Day, San Francisco, 1970.
4. A. Halati. *Estimation of the Ratio of the Steady-State Means of a Bivariate Stochastic Process.* Ph. D. thesis. University of Southern California, Los Angeles. Jan. 1985.
5. E. C. Fieller. "Some Problems in Interval Estimation." In *Journal of the Royal Statistical Society*, Series B, No. 2, 1954, pp. 175–185.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

A Model for Predicting Free-Flow Speeds Based on Probabilistic Limiting Velocity Concepts: Theory and Estimation

THAWAT WATANATADA AND ASHOK M. DHARESHWAR

Methods for predicting space-mean speeds over a heterogeneous roadway for a freely moving vehicle are needed for many applications in highway planning. Although various shortcomings of a linear specification for the steady-state speed prediction model have been recognized and alternative models based on a limiting speed specification have been proposed, the latter have not previously been rigorously estimated. Presented are the theoretical formulation and empirical estimation of a new probabilistic model for predicting steady-state speeds on a homogeneous road section, based on a limiting velocity approach. The statistical implementation employing a multinomial logit formulation and estimation results using a Brazilian data set in excess of 100,000 observations are presented. Alternative methods have been developed for applying the steady-state speed model to predict speeds over heterogeneous roadways under different informational limitations and accuracy requirements. The speed model, together with related models for predicting fuel consumption and tire wear, have been incorporated in the World Bank Highway Design and Maintenance Standards Model (HDM-III) to provide a basis for engineering-economic analysis of alternative standards of geometric design and pavement design and maintenance for low-volume roads.

A method for predicting the costs of operating a vehicle on a highway of known characteristics is an important tool for highway sector planning and project evaluation. In general, a method for predicting vehicle operating costs per unit roadway distance consists of *(a)* a central model to predict speed and related variables (e.g., power used); and *(b)* a set of interfacing models that would use the predictions from the central model as inputs, and generate predictions of journey time, fuel consumption, tire wear, and vehicle utilization *(1)*.

Some of the desirable properties for the central component, the speed prediction model, are as follows:

- It should be flexible in its input requirements;
- It should be appropriately sensitive to the policy options being evaluated (e.g., design parameters and road maintenance resource expenditure);
- It should be amenable to extrapolation over a reasonable range of the policy variables; and
- It should be readily transferable to other environments.

The purpose of the paper is to describe the limiting speed approach to predicting steady-state speeds of vehicles on a homogeneous road section and its statistical implementation leading to the formulation of a probabilistic limiting speed model. Estimation results using a large data set of speed observations collected in Brazil are presented and discussed.

T. Watanatada and A. M. Dhareshwar, The World Bank, 1818 H St., N.W., Washington, D.C. 20433.

The probabilistic limiting speed approach to modeling vehicle operation on a roadway bears a close similarity to the random utility approach to modeling urban travel demand among discrete alternatives. This similarity has been exploited in resolving some of the issues relating to model estimation, aggregation, and transferability.

The steady-state speed prediction model is the centerpiece of the vehicle operating cost module of the World Bank Highway Design and Maintenance Standards Model—Release 3 (HDM-III) *(1–3)*. The linkages between the speed model and various vehicle cost components are indicated in the concluding section.

STEADY-STATE SPEED PREDICTION MODEL

To cope with the diversity of input information availability and output accuracy requirements that the speed prediction model is called on to cope with, it is convenient to structure the process of speed prediction into two components:

1. A model for predicting the vehicle speed on a road section over which the characteristics of interest do not change appreciably. Such a section is referred to as a homogeneous section. The concept of speed used is that of a steady-state speed, and this component of speed prediction may be called a steady-state speed prediction model.
2. A set of procedures with which to apply the steady-state speed model for predicting the speed profile over a heterogeneous roadway by using the available information on the roadway and with the desired degree of accuracy for the particular application. These procedures are called roadway speed prediction methods.

The steady-state speed of an unimpeded vehicle of known attributes traversing a homogeneous road section of known characteristics, located in a fixed overall socioeconomic and traffic environment, may be defined as the speed the vehicle would eventually attain and maintain if the homogeneous road section is indefinitely long. Thus steady-state speed in a fixed environment is a property associated with a given combination of a homogeneous road section and a vehicle.

A homogeneous road section is assumed to be completely defined if its surface type, slope, curvature, superelevation, and surface irregularity measure are specified. The section is assumed to be sufficiently wide so that the road width has no effect on the speed—as was the case with roads used in speed observation in Brazil. It may be observed that the highway

characteristics represent the policy variables under the control of the highway planner. These are occasionally referred to as speed-influencing characteristics or road severity factors collectively denoted by the symbol X. It may be noted that the direction of travel on the homogeneous section is part of the description of the section.

The term vehicle is shorthand for operator-vehicle system and the term operator is used to indicate that the speed decision maker may be an individual driver or a transport firm. As for the characteristics of the vehicle, the vehicle class and loading (in the case of a truck) are supposed to be known. Also, a set of technical characteristics of the vehicle—such as unladen weight, drag coefficient, and so on—are assumed to be known or assignable with reasonable accuracy. The technical characteristics of the vehicle are denoted by the symbol Y.

Finally, there is a set of behavioral-technical characteristics of the vehicle, such as used power, perceived friction ratio, desired speed, and so on. These are the estimated parameters of the steady-state speed prediction model and are collectively denoted by the symbol θ. For a given application these parameters may be estimated afresh, calibrated on the basis of limited observations or, in some cases, judgmentally determined.

A specification of a steady-state speed prediction model is a functional form relating the steady-state speed, V (in m/s), of the given vehicle to variables that capture various speed-influencing characteristics of the homogeneous section. The variables considered are given in Table 1.

TABLE 1 VARIABLES CONSIDERED

Speed-Influencing Characteristic	Variable Name	Units	Symbol	Note
Vertical alignment	Gradient or slope	Fraction	GR	With sign
Horizontal alignment	Curvature	rad/km	C	Without orientation
	Superelevation	Fraction	SP	
Surface irregularity	Roughness	m/km IRI	R	See note below
Surface type	–	–	ST	Paved/unpaved

Note: IRI is the International Roughness Index.

The unit of roughness used is the International Roughness Index (IRI), which summarizes the varied wavelengths and amplitudes of the surface irregularities in a slope index that is equivalent to the total axle-body movement (in meters) made by a typical passenger car over a unit distance (in kilometers). The index quantifies the impact of roughness on a moving vehicle in much the same way as roughness causes vehicle costs, and as such is judged to be the most applicable measure of roughness for economic evaluation purposes *(4,5)*.

The models found in the literature may generally be classified into two approaches: direct and latent (or unobservable) variable approaches. The basic specification using the direct approach would be the linear form

$$V = a_0 + a_1GR + a_2C + a_3R \quad (1)$$

where

a_0, a_1, a_2, and a_3 = parameters to be estimated,
GR = gradient or slope (fraction),
C = curvature (rad/km), and
R = roughness (m/km IRI).

The direct approach is not entirely satisfactory for the following reasons:

- It is possible to predict unreasonably low (at times negative) steady-state speeds while using plausible values for the independent variables, especially for low standard roads where these independent variables assume large values.
- The partial derivatives of predicted steady-state speed with respect to each of the road severity factors is constant and, as just noted, in general, negative. The policy implication of this property of linear specifications is that reduction of one of the road severity factors (such as roughness), through a greater investment in road maintenance, will show an increase in speeds, although another factor (such as gradient) may actually be inhibiting speeds.

Although it is possible partially to mitigate these shortcomings by making the functional form nonlinear and by including interaction terms, the approach would still be ad hoc.

DETERMINISTIC LIMITING SPEED MODEL

The alternative approach for specifying a steady-state speed model is to use a set of limiting speeds or constraining speeds as latent or unobservable variables. Instead of associating the steady-state speed directly with the speed-influencing variables X of the homogeneous road sections, these variables are regarded as interacting with the relevant characteristics of the vehicle Y to generate a set of steady-state speed constraints. The resulting steady-state speed of the vehicle on the homogeneous section is then postulated to be the maximum attainable speed subject to these constraints. In symbols,

$$V^\circ = \min(\bar{V}_g, \bar{V}_c, \bar{V}_r, \bar{V}_d) \quad (2)$$

where

V° = deterministic steady-state speed;
$\bar{V}_g$ = gradient-limited speed,
$\bar{V}_c$ = curvature-limited speed,
$\bar{V}_r$ = roughness-limited speed, and
$\bar{V}_d$ = desired speed in the absence of road severity factors,

All of the speeds are expressed in meters per second.

Figure 1 shows the resultant steady-state speed, V°, and the limiting speeds as functions of gradient for a laden heavy truck traversing slightly curvy and relatively smooth homogeneous sections. As an example of interpreting the plot, if the vehicle is traveling uphill on a homogeneous section with 1 percent gradient, the values of limiting speeds are as follows: $\bar{V}_g$ = 18 m/sec, $\bar{V}_c$ = 68 m/sec, $\bar{V}_r$ = 52 m/sec, and $\bar{V}_d$ = 24 m/sec.

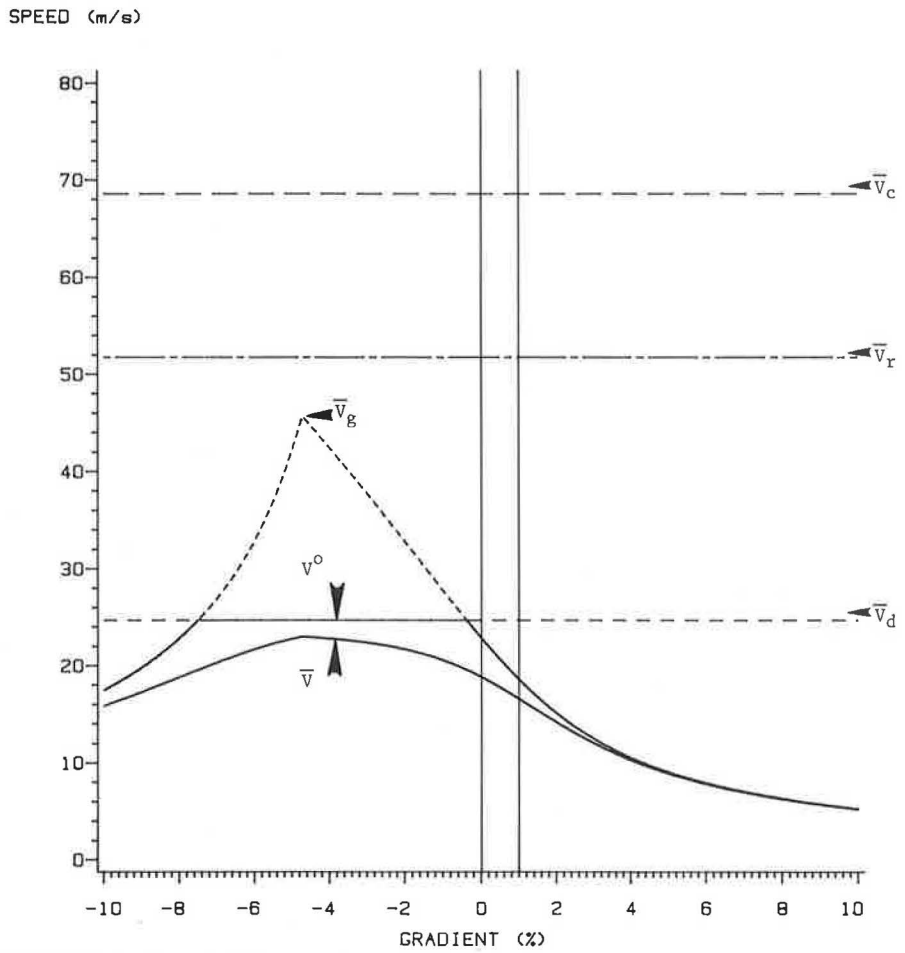


FIGURE 1 Constraining speeds and the steady-state speeds as functions of gradient.

Hence, the steady-state speed V° is 18 m/s.

The steady-state speed curve has three distinct regimes. For very steep negative gradients (of magnitude greater than 7.5 percent), the steady-state speed is governed by the upward-sloping part of the gradient-limited speed, $\bar{V}_g$, which depends on the braking power used. Over the middle of the gradient range, the steady-state speed is determined by the desired speed, $\bar{V}_d$. Finally, for slightly negative grades and for all positive grades, the downward-sloping part of $\bar{V}_g$, which depends on the used driving power, dominates. In this example, the curvature-limited speed, $\bar{V}_c$, and the roughness-limited speed, $\bar{V}_r$, do not have a discernible influence on the steady-state speed.

Thus, the steady-state speeds predicted by using a limiting velocity formulation show asymptotically consistent behavior. That is, as various severity factors deteriorate from their ideal values, the predicted speeds decrease monotonically but retain plausible values. Further, as will be shown, a considerable amount of scientific, technological, and behavioral information can be incorporated in relating the limiting speeds to road section and vehicle characteristics.

The limiting speed approach to steady-state speed prediction has been used in a number of studies *(6–10)*. The studies differ in the number of limiting speeds used as well as in the way they are related to road and vehicle characteristics. The derivation of the constraining speeds used in this study is described next.

Gradient-Limited Speed, $\bar{V}_g$

The limiting speed governed by the vertical alignment of the homogeneous road section is derived based on two considerations, namely, the driving power used and the braking power used, thus giving rise to two basic limiting speeds, V_{dr} and V_{br}, respectively. $\bar{V}_g$ is taken to be the lower of these two speed constraints. That is,

$$\bar{V}_g = \min(V_{dr}, V_{br}) \tag{3}$$

where V_{dr} is the speed governed by driving power and gradient (m/sec), and V_{br} is the speed governed by braking power and gradient (m/sec).

V_{dr} and V_{br} are derived by first making behavioral assumptions about the use of driving and braking power, respectively, and then relating speeds to powers and gradients through the force-balance relation. Under steady-state conditions, the force balance may be written as

$$1{,}000P/v = mg(GR + CR) + 0.5\rho c_d a v^2 \tag{4}$$

where

P = used power (kW);
v = speed (m/sec);

m = vehicle mass (kg), which is $m_0 + m_1$, where m_0 = mass of the empty vehicle (kg) and m_1 = net load (kg);
g = gradient;
GR = section gradient (as a fraction and with sign);
CR = coefficient of rolling resistance (dimensionless);
ρ = mass-density of air (kg/m^3) given by $1.225(1 - 226h10^{-5})^{4.225}$, where h is the elevation of the section over the mean sea level (m);
c_d = drag coefficient of the vehicle (dimensionless); and
a = projected frontal area of the vehicle (m^2).

SPEED GOVERNED BY DRIVING POWER, V_{dr}

V_{dr}, the speed limited by driving power used and gradient, is arrived at based on the assumption that when vertical gradient is the only road severity factor, the vehicle is driven at steady-state speed using a constant level of driving power. Denoting the constant driving power used by PDRIVE (in kW) and substituting PDRIVE for P in the force balance (Equation 4), yields

$$0.5\rho c_d a v^3 + mg(GR + CR)v - 1{,}000\ \text{PDRIVE} = 0 \quad (5)$$

which is a cubic equation in the unknown quantity v. It may be observed that for all values of GR, the number of sign changes in the coefficients of the equation is one and hence, by Descartes's rule of signs, the equation always has exactly one positive root *(11)*. V_{dr} is defined as the unique positive solution to Equation 5.

Cubic equations are generally solved iteratively. However, because the coefficient of v^2 in Equation 5 is zero, the equation has a relatively tractable analytical solution, which is given by Watanatada et al. *(1)*.

Speed Governed by Braking Power, V_{br}

V_{br}, the limiting speed determined by braking power used and gradient, is arrived at on the basis of the postulate that when a vehicle descends a long steep grade its descent speed is controlled by the vehicle braking capability, which results from the use of the vehicle engine retardation power or the regular brakes, or both. Thus, V_{br} is analogous to the concept of braking crawl speed. It is assumed that when negative gradient is the only road severity constraint, the steady-state speed is attained by using a constant level of braking power, a positive quantity by convention denoted by PBRAKE (in kW). Substituting P = –PBRAKE in the force balance (Equation 4),

$$0.5\rho c_d a v^3 + mg(GR + CR)v + 1{,}000\ \text{PBRAKE} = 0 \quad (6)$$

which is again a cubic equation in the unknown speed v. When the value of the effective gradient (i.e., $GR + CR$) is negative, the equation has two distinct positive roots; it is not, in general, possible to identify the physically meaningful solution on a priori grounds. However, because the braking speed constraint is likely to become binding only on steep negative grades where the steady-state speeds will be relatively low, the contribution of air resistance to the force balance may be neglected without serious error. Further, it is expected that the braking speed constraint be inapplicable when positive power is needed to move the vehicle, that is, when the effective gradient is non-negative. Thus, the limiting speed due to gradient and braking power may be obtained as

$$V_{br} = \begin{cases} \infty & \text{if } GR + CR \geq 0 \\ -1{,}000\ \text{PBRAKE}/[mg(GR + CR)] & \text{if } GR + CR < 0 \end{cases} \quad (7)$$

Figure 2 shows the V_{dr}, V_{br}, and $\bar{V}_g$ curves as functions of gradient for a laden heavy truck on a homogeneous section (with a rolling resistance coefficient of 0.015.).

For the coefficient of rolling resistance, the following relationship expressing it as a function of section roughness, estimated in the Brazil study *(1)*, may be used

$$CR = \begin{cases} 0.0139 + 0.00026R & \text{for buses and trucks} \\ 0.0218 + 0.00061R & \text{for cars and utilities} \end{cases} \quad (8)$$

Curvature-Limited Speed, $\bar{V}_c$

The limiting speed governed by curvature of the homogeneous road section is arrived at from the postulate that when curvature is significant, the speed is limited by the tendency of the wheels to skid. An appropriate measure of the tendency to skid is the ratio of the lateral force on the vehicle to the normal force, which may be termed the perceived friction ratio. Under the assumption that when curvature is the only constraining road severity factor, the steady-state speed of a vehicle is attained by using a constant perceived friction ratio, denoted by FRATIO, an expression for $\bar{V}_c$ may be derived as follows.

The lateral force (LF) and the normal force (NF) on the vehicle are (in newtons)

$$LF = mv^2(C/1{,}000)\cos SP - mg \sin SP \simeq mv^2(C/1{,}000) - mgSP \quad (9)$$

and

$$NF = mg \cos SP + mv^2(C/1{,}000) \sin SP \simeq mg + mv^2(C/1{,}000)SP \quad (10)$$

where C is the section curvature (rad/km), and SP is the section superelevation (expressed as a fraction).

Thus

$$\text{FRATIO} = \text{LF/NF} = [(v^2C/g\ 1{,}000) - SP]/[1 + SPv^2C/(g\ 1{,}000)] \simeq v^2C/(g\ 1{,}000) - SP$$

Taking $\bar{V}_c$ to be the positive root of the above quadratic in v,

$$\bar{V}_c = [(\text{FRATIO} + SP)\ g1{,}000/C]^{1/2} \quad (11)$$

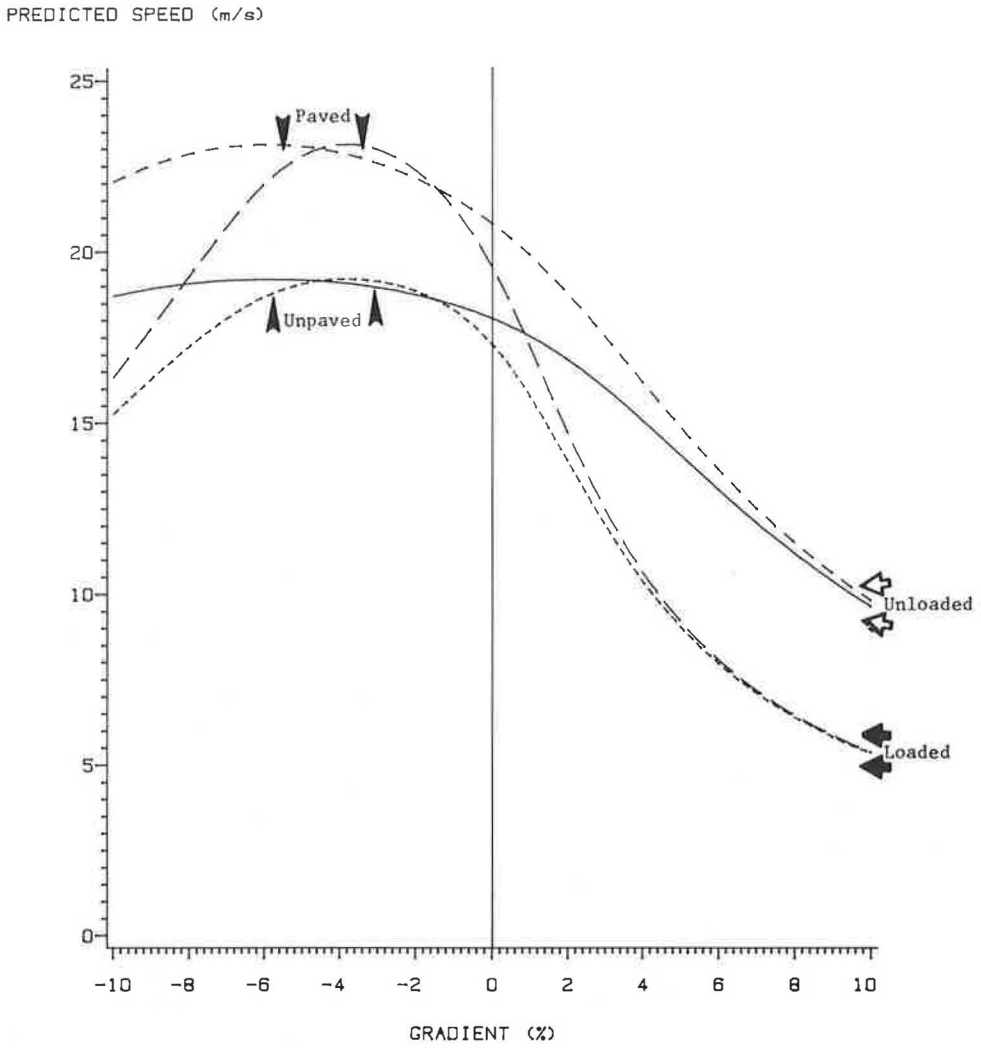


FIGURE 2 Predicted speed of a heavy truck as a function of gradient.

The parameter is called the perceived friction ratio to distinguish it from the actual friction ratio, which is the ratio of the vectorial sum of the lateral and drive forces on the vehicle to the normal force. The curvature-limited speed can also be modeled based on lateral acceleration or sight distance considerations.

From the Brazil data, the FRATIO parameter for a vehicle class has been found to depend on the surface type and, for trucks, the loading condition.

Roughness-Limited Speed, $\bar{V}_r$

This speed constraint is derived based on the notion that when roughness is the only prevailing road severity factor, the vehicle speed is limited by the discomfort sensed or the severity of the ride. An adequate measure of the ride severity for a vehicle with a rigid rear axle is the average rectified velocity (ARV) (in mm/s), defined as the rate of cumulative absolute displacement of the rear-axle relative to the vehicle body. It is approximately proportional to vehicle speed and road roughness, and may be written as

$$\text{ARV} = \text{ARV}(v) = 1.15vR \tag{12}$$

where R is the section roughness (in m/km IRI), and the constant of proportionality reflects a calibration factor for the Maysmeter-equipped Opala automobile used in the Brazil study and unit conversion factors *(12, 13)*.

When roughness is the only constraining road severity factor, the steady-state speed for a vehicle is assumed to be attained at a constant representative value of average rectified velocity, denoted by ARVREP (in mm/s); under this assumption an expression for $\bar{V}_r$ is achieved by solving

$$\text{ARVREP} = 1.15vR$$

That is,

$$\bar{V}_r = \text{ARVREP}/(1.15R) \tag{13}$$

It will be seen that, for the Brazil data, the ARVREP parameter does not vary significantly across surface types and load classes, only over the vehicle classes. Thus, given the roughness of the homogeneous section, the roughness-constrained

speed, $\bar{V}_r$, may be computed if an estimate of the ARVREP parameter is available.

Desired Speed, $\bar{V}_d$

Finally, $\bar{V}_d$ is the desired speed, that is, the speed at which a vehicle of a given class would be operated in the absence of constraints based on gradient, curvature, and roughness. The desired speed results from the driver's response to psychological, safety, economic, and other considerations (for example, speed limits or even driver's perception of the strictness of enforcement), and, as such, it can be related to a number of factors. (In an extension of the current model using Indian data, it depends on the width class of the homogeneous section.) In the current model, $\bar{V}_d$ has been assumed to be constant for a given surface class, and estimated directly as a model parameter.

PROBABILISTIC LIMITING SPEED MODEL: FORMULATION

Even if the assumption of constant θ were true, the limiting speeds would still vary over different homogeneous sections and over different vehicles of the same class; these variations could be only partially explained by the variation in the observed characteristics of the section *(X)* and vehicle *(Y)*. Some of the important reasons are measurement errors, omission of characteristics of the road section and vehicle, deviations of the characteristics X and Y from the values actually used, the inability of the observer to determine the binding constraint with certainty, and the inability of the modeler to completely specify the decision procedure of the vehicle operator. In sum, the limiting speeds have to be treated as random variables and the parameters have to be estimated on this basis. It is the explicit recognition of the stochastic nature of the constraining speeds that distinguishes the probabilistic steady-state speed prediction model presented here from those of the earlier studies.

This notion is formalized by treating the limiting speeds as random variables (or variates) with means or expected values given by the expressions derived in the deterministic version. For example, denoting the gradient-limited speed variate by $V_g = V_g(X,Y{:}\theta)$, it may be written that

$$V_g(X, Y{:}\ \theta) = \bar{V}_g(X, Y{:}\ \theta)\eta_g(X, Y{:}\ \theta)$$

or, suppressing the arguments for simplicity,

$$V_g = \bar{V}_g\eta_g$$

Treating the other limiting speeds analogously,

$$V_z = \bar{V}_z\eta_z, \quad \text{for } z = g, c, r, \text{ and } d \tag{14}$$

Next, a new random variable V is defined as

$$V = \min (V_g, V_c, V_r, V_d) \tag{15}$$

Just as the speed constraint random variables, V may be expressed as

$$V(X,Y{:}\ \theta) = \bar{V}(X, Y{:}\ \theta)\eta(X, Y{:}\ \theta) \text{ or } V = \bar{V}\eta \tag{16}$$

It should be noted that although

$$V = \min (V_g, V_c, V_r, V_d)$$

it is generally not the case that

$$\bar{V} = \min (\bar{V}_g, \bar{V}_c, \bar{V}_r, \bar{V}_d)$$

In fact,

$$\bar{V} \leq \min (\bar{V}_g, \bar{V}_c, \bar{V}_r, \bar{V}_d)$$

with the equality holding if and only if the random variables are perfectly positively correlated or they are degenerate, that is, the variations are all zero. In other words, the mean of the minimum is generally less than the minimum of the means.

Thus the relation between the means of the speed constraints and the mean steady-state speed depends on the assumptions imposed on the joint distribution of the errors of the speed constraint variates. The error structure is specified and the estimation is performed by making use of two well-known distributions (lognormal and normal) and two distributions from a class known as asymptotic extreme value distributions. These are the Weibull distribution and the Gumbel distribution *(14–16)*.

Just as normal distributions are preserved when the arithmetic operation involved is one of addition, the Weibull and the Gumbel distributions are preserved when the arithmetic operation involved is minimization (or maximization). This property enables one to derive the distribution of the minimum variate as a closed form function.

The disturbances pertaining to a particular speed observation and the associated speed constraints are specified by using three nested components of error. First, there are errors $\varepsilon(X)$, pertaining to the homogeneous section, which include unmeasured characteristics of the section and speed measurement errors. Second, there are errors $\zeta(X,Y)$, pertaining to the particular vehicle observed at that section, which include unmeasured characteristics of the particular vehicle at the section. Finally, given these two errors, there would be errors $\tau_z(X,Y)$, specific to the various speed constraint variates for that speed observation. That is, with η_z as the random part for a given realization of a constraining speed variate v_z for vehicle Y on section X,

$$v_z = \bar{V}_z\eta_z = \bar{V}_z[\varepsilon(X)][\zeta(X, Y)][\tau_z(X, Y)] \tag{17}$$

Proceed by imposing fairly standard assumptions of lognormality regarding the first two components of error. The Weibull distribution will be used for the third component to derive the conditional distribution of the observed speed variate. Specifically,

- Errors $\varepsilon(X)$ are independent and have identical lognormal distributions with mean 1.

- Errors $\zeta(X,Y)$ are independent and have identical lognormal distributions with mean 1.
- Errors $\tau_z(X,Y)$ are independent and have identical Weibull distributions with mean 1 and shape parameter β.

Under these three assumptions, by using the properties of the Weibull distribution, the following results are obtained:

- Conditional on $\varepsilon(X)$ and $\zeta(X,Y)$, the attained speed is a variate V, which has a Weibull distribution with a shape parameter β.
- The relationship between the conditional means of the attained speed and the limiting speed variates is

$$\bar{V} = (\bar{V}_g^{-1/\beta} + \bar{V}_c^{-1/\beta} + \bar{V}_r^{-1/\beta} + \bar{V}_d^{-1/\beta})^{-\beta} \quad (18)$$

where $\bar{V}_x = \bar{V}_x(X,Y: \theta)$ and hence, $\bar{V} = \bar{V}(X,Y: \theta, \beta)$.

Thus, the speed observation may be written as a random variable V, with

$$V = \bar{V}\eta = \bar{V}[\varepsilon(X)][\zeta(X, Y)][\tau(X, Y)]$$

where $\tau(X,Y)$ have independent Weibull distributions with mean 1 and shape parameter β, or

$$V = \bar{V}\varepsilon\omega \quad (19)$$

where $\omega = \omega(X,Y) = [\zeta(X,Y)][\tau(X,Y)]$ which is approximately lognormal.

Equation 18, along with the expressions for $\bar{V}_x$ given earlier, constitutes a multinomial logit model that is nonlinear in the parameters θ and β *(17–19)*. Figure 1 shows $\bar{V}$ as a function of gradient. At 1 percent gradient, the predicted speed is approximately 16 m/s.

Equivalently, the model can be expressed in terms of the logarithms of speeds. This version is more convenient for estimation purposes. Assuming that for a given speed observation the logarithms of the constraining speeds have independent Gumbel distributions with identical scale parameters β, the properties of Gumbel distribution can be used to express the model as follows. Defining

$$U = \ln V \quad (20)$$

$$U_z(X, Y: \theta) = \ln V_z(X, Y: \theta), \text{ for } z = g, c, r, \text{ and } d \quad (21)$$

leads to

$$\bar{U} = \ln [\exp(-\bar{U}_g/\beta) + \exp(-\bar{U}_c/\beta) + \exp(-\bar{U}_r/\beta) + (-\bar{U}d/\beta)]^{-\beta} \quad (22)$$

where $\bar{U}_z = \bar{U}_z(X,Y: \theta)$ are the means of the random variables $\bar{U}_z$. Thus by analogy with the speed model, the log speed model may be written as

$$U = \bar{U} + e + w \quad (23)$$

where $e = e(X)$ have independent normal distributions with constant mean and variance, and $\omega = w(X,Y)$ are independent and identically distributed as convolutions of independent normal and Gumbel distributions. The distribution of w is approximately normal.

The estimation problem then is to find estimates θ and β that satisfy a suitable criterion. The two most commonly used criteria for estimating a logit model are maximum likelihood and least squares *(20–22)*. The criterion chosen was least squares.

PROBABILISTIC LIMITING VELOCITY MODEL: ESTIMATION

Data for Estimation

The data for the steady-state speed model were obtained from radar speed observations of vehicles at selected homogeneous road sections over a period of about 1 year. Sections were distinguished by direction of travel. Because the roughness of these sections varied significantly over nonconsecutive observation periods, the unpaved sections were further distinguished by roughness intervals of 4 m/km IRI. This procedure resulted in a total of 216 homogeneous sections with gradient ranging from –9 to 11 percent, curvature from 0 to 50 rad/km, and roughness from 1.5 to 15 m/km IRI.

The observations made for each vehicle sighting were spot speeds on the section, vehicle type, and load condition. The vehicle types observed were categorized into six classes, and the three truck classes were further divided into unloaded and loaded categories. These classes are given in Table 2 with the adopted average vehicle characteristics. The average gross vehicle masses were obtained from a separate axle load study for Brazil. The values of aerodynamic drag coefficient and frontal area were adapted from those for typical makes and models prevalent in Brazil for each vehicle class.

TABLE 2 VEHICLE CLASSES AND CHARACTERISTICS USED IN STEADY-STATE SPEED MODEL ESTIMATION

Vehicle Class	Drag Coefficient, c_d	Projected Frontal Area, a (m^2)	Total Vehicle Mass, m, (kg)
Car	0.50	2.00	1 200
Utility	0.60	3.00	2 000
Bus	0.65	6.30	10 400
Light/medium truck	0.70	4.5	5 400 (unloaded) 11 900 (loaded)
Heavy truck	0.85	5.2	7 900 (unloaded) 19 200 (loaded)
Articulated truck	0.65	5.8	15 900 (unloaded) 37 700 (loaded)

In all, about 100,000 speed observations were included. For each vehicle class, the logarithms of individual speed observations pertaining to a section were averaged, yielding the dependent variable values of the estimation data set.

TABLE 3 ESTIMATION RESULTS OF STEADY-STATE SPEED MODEL FOR SIX VEHICLE CLASSES: ESTIMATES

Vehicle Class	β (a)	PDRIVE, kW (b)	PBRAKE, kW (c)	ARVREP, mm/s (d)	V m/s: Unpaved Surface (e)	V m/s: Paved Surface Increment over Unpaved Value (f)	V m/s: Paved Surface Value (e) + (f)	FRATIO: Unpaved Surface (unloaded or loaded vehicle) (g)	FRATIO: Paved Surface, Unloaded Vehicle Increment over Unpaved Value (h)	FRATIO: Paved Surface, Unloaded Vehicle Value (g) + (h)	FRATIO: Paved Surface, Loaded Vehicle Increment over Paved Unloaded (i)	FRATIO: Paved Surface, Loaded Vehicle Value (g) + (h) + (i)
Car	0.274	26.8	16.0	259.7	22.8	4.5	27.3	0.124	0.144	0.268	–	–
Utility	0.306	32.7	24.0	239.7	21.8	4.6	26.4	0.117	0.104	0.221	–	–
Bus	0.273	83.1	157.3	212.8	19.3	6.7	26.0	0.095	0.138	0.233	–	–
Light/medium truck	0.304	69.7	140.4	194.0	20.0	2.7	22.7	0.099	0.154	0.253	–0.083	0.170
Heavy truck	0.310	79.6	189.2	177.7	20.0	4.7	24.7	0.087	0.205	0.292	–0.107	0.185
Articulated truck	0.244	147.2	368.0[a]	130.9	13.8	9.6	23.4	0.040[a]	0.139	0.179	–0.049	0.130

[a]This parameter for the articulated truck class was exogenously assigned.

Estimation Results

The final results are presented in Table 3. These results consist of six sets of parameter estimates, for cars, utilities, buses, light and medium trucks, heavy trucks, and articulated trucks. The asymptotic *t*-statistics associated with these parameter estimates are given in Table 4 in a similar format. A goodness-of-fit measure analogous to the R^2 value in linear models is also given. This was obtained by regressing the mean observed speeds against predicted speeds.

There were too few observations for articulated trucks to support the determination of all the model parameters and the PBRAKE and FRATIO parameters were assigned the values of 368 kW and 0.40, respectively, when the estimation was carried out.

All the parameter estimates have the expected sign and most have the expected relative magnitudes as individually discussed in the following paragraphs. All but one of the parameter estimates have asymptotic *t*-statistics significant at approximately 5 percent.

Except for the mixed class of light and medium trucks, the magnitudes of the driving power used (PDRIVE) are consistently smaller than the maximum rated power values of the typical vehicles of the respective classes. In fact, there is an approximate relationship between these quantities that can be used to calibrate the PDRIVE parameter for a new vehicle *(1)*.

The magnitudes of the braking power used (PBRAKE) appear to increase with the gross vehicle mass. As would be expected, the greater the mass of a vehicle, the more the braking capability needed to render the vehicle operations safe. An approximate relationship between the braking power used and the gross vehicle mass may also be derived.

The FRATIO estimates, from 0.087 to 0.292, are well below the range of 0.6 to 0.7 found from skid-pad tests of modern high-performance passenger cars. This appears to indicate a large margin of safety within which vehicles are generally

TABLE 4 SOME IMPORTANT STATISTICS ASSOCIATED WITH THE ESTIMATION

Vehicle Class	β (a)	PDRIVE (b)	PBRAKE (c)	ARVREP (d)	V_d: Unpaved Surface (e)	V_d: Increment for Paved Surface (f)	FRATIO: Unpaved Surface (g)	FRATIO: Increment for Paved Unloaded (h)	FRATIO: Further Increment for Paved Loaded (i)	Variances σ_e^2	Variances σ_w^2	R^2	No. of Observations	Sum of Squared Residuals
Car	10.9	15.2	7.9	20.3	22.8	4.9	12.4	10.3	–	0.00654	0.0224	0.92	216	1.36
Utility	10.9	19.3	7.9	17.5	17.4	4.0	9.3	7.0	–	0.00808	0.0355	0.89	216	1.68
Bus	6.8	27.3	4.6	12.0	12.0	3.8	5.3	5.3	–	0.02477	0.0276	0.83	216	5.15
Light/medium truck	13.4	44.0	9.5	24.6	18.4	2.8	9.0	7.3	–3.7	0.01574	0.0405	0.87	431	6.64
Heavy truck	9.7	41.6	10.9	19.1	11.6	2.8	3.8	5.9	–3.1	0.02578	0.0369	0.85	381	9.59
Articulated truck	7.0	33.9	–	19.0	11.6	7.0	–	5.0	–1.5	0.03588	0.0365	0.81	232	8.07

Note: The members in the parameter columns are the respective asymptotic *t*-statistics.

operated on public roads. Within a vehicle class, the estimates for unpaved roads are significantly smaller than those for paved roads. Further, for paved road operations, laden trucks have smaller FRATIO estimates than unladen trucks. Finally, across the vehicle classes, the FRATIO estimates tend to vary inversely with the size of the vehicle.

The estimates of the average rectified velocity (ARVREP) show a clear tendency to vary inversely with the vehicle size, the cars having the largest value and the articulated trucks the smallest. This is somewhat surprising because on purely physical reasoning it would be expected that the smaller vehicles be more sensitive to road roughness than the larger ones. The reversal of relative magnitudes is probably explained in part by the higher tire stiffness of larger vehicles and in part by the economic response of the driver to the relatively higher cost impact of roughness on larger vehicles.

The estimates for the desired speed ($\bar{V}_d$) are, as expected, higher for paved roads than for unpaved roads. Moreover, they tend to be larger for smaller vehicles, although they are relatively constant for each surface type for most vehicle classes.

The discriminating power of the models may be seen in Figures 2–4, which show graphs of predicted steady-state speed plotted against the gradient, curvature, and roughness, respectively—for unloaded and loaded heavy trucks—for both paved and unpaved surfaces.

APPLICATIONS AND DIRECTIONS FOR FUTURE RESEARCH

The most important current application of the steady-state speed prediction model has been the development of three methods for predicting speeds on heterogeneous roadways.

1. The micro-transitional roadway speed prediction method, which uses detailed information on the roadway and simulates transitional driver behavior including speed change cycles.
2. The micro-nontransitional method, which requires the characteristics of all the homogeneous sections of the roadway, but does not model transitional driver behavior.
3. The aggregate method, which uses a classification-based aggregation procedure, which is fairly widely used in demand aggregation *(18, 23)*.The method uses summary descriptors of the characteristics of the roadway and generates speed predictions for one-way and round-trip travels on the roadway. The aggregate method, which incorporates the effect of road width

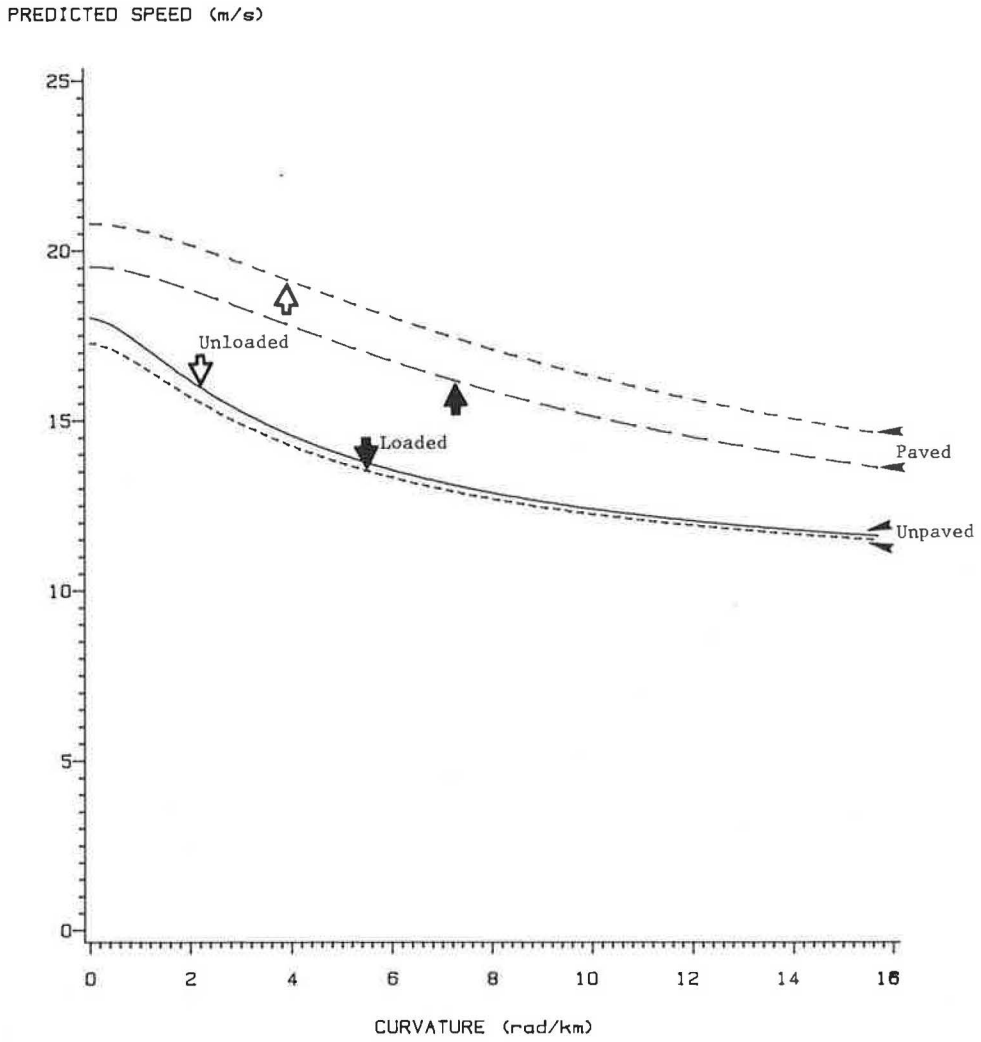


FIGURE 3 Predicted speed of a heavy truck as a function of curvature.

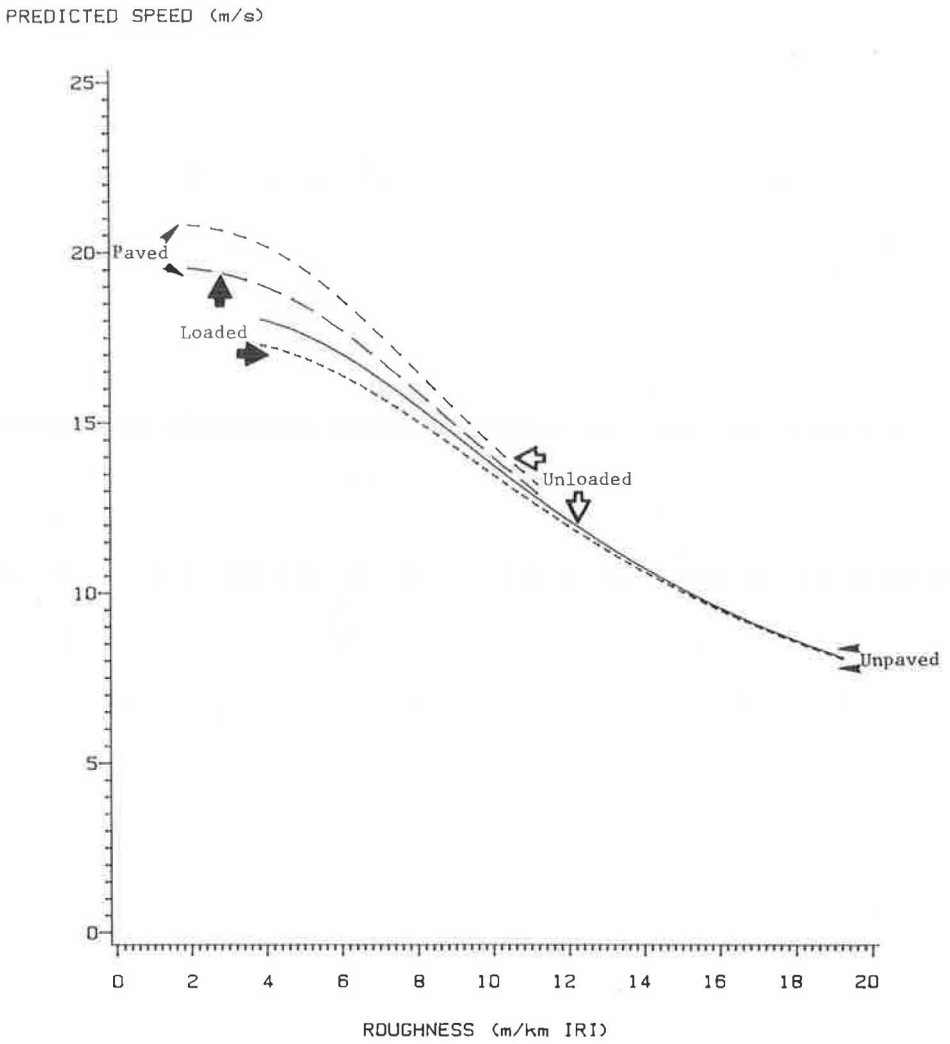


FIGURE 4 Predicted speed of a heavy truck as a function of roughness.

based on data from India, has been selected for use in the HDM-III model *(2)*.

Models have been developed that use predicted speeds and other derivative variables (such as power used, tangential energy, block speed, etc.) as inputs to predict fuel consumption, tire wear, and vehicle utilization. Depending on the method of speed prediction used, these models can provide predictions on a disaggregate or aggregate basis. The aggregate versions of these models have also been implemented in HDM-III.

A methodology to calibrate the parameters of the steady-state speed prediction model for a new environment without full-fledged estimation has been developed based on a nonrandom sampling technique *(2, 24)*.

Potentially the most important area for future application of the speed prediction method is in modeling vehicle interaction and the consequences for vehicle operating costs, both at the disaggregate and aggregate levels.

Possible improvements to the steady-state speed prediction model include the following:

- The relationships between the speed constraints and speed-influencing characteristics of the road section could be further enhanced by taking into account the effect of the following: shoulder width and condition on the desired speed, curvature on the braking capacity, and lateral acceleration and sight distance on perceived friction ratio.
- In regard to the distributional assumptions made in the probabilistic version of the model, the most restrictive ones are independence and equality of the shape parameter for all the speed constraint variates. It would be of interest to test the acceptability of these assumptions by estimating the model parameters under the more general multinomial probit formulation.

ACKNOWLEDGMENTS

Presented in this paper is a part of the large array of results emanating from the Highway Design and Maintenance Standards study (HDMS) of the Transportation Department of The World Bank. The field data on which the findings presented here are based were assembled by the highway research project, PICR, sponsored by the government of Brazil, the United Nations Development Programme, and The World Bank. The project was executed by GEIPOT, the Brazilian transportation

agency, in collaboration with the Texas Research and Development Foundation as the agents of The World Bank. The work, being part of as large an undertaking as HDMS, has received contributions from many individuals from these organizations. Because they are too numerous to be mentioned in the space available, the authors wish to take this opportunity to acknowledge them collectively. The authors are also grateful to the three anonymous reviewers for helpful comments.

REFERENCES

1. T. Watanatada et al. *The Highway Design and Maintenance Standards Model (HDM-III)*, Vol. 2: Models for Predicting Vehicle Operating Costs Based on Mechanistic Principles: Theory and Quantification. Transportation Department, The World Bank, Washington, D.C., 1985,
2. T. Watanatada et al. *The Highway Design and Maintenance Standards Model (HDM-III)*, Vol. 4: Model Description and User's Manual. Transportation Department, The World Bank, Washington, D.C., 1985.
3. A. Chesher and R. Harrison. *The Highway Design and Maintenance Standards Model (HDM-III)*, Vol. 1: User Benefits from Highway Improvements: Evidence from Developing Countries. Transportation Department, The World Bank, Washington, D.C., 1985.
4. W. D. O. Paterson. *Highway Design and Maintenance Standards Model (HDM-III)*, Vol. 3: Prediction of Road Deterioration and Maintenance Effects: Theory and Quantification. Transportation Department, The World Bank, Washington, D.C., 1985.
5. W. D. O. Paterson. Relationship of the International Roughness Index to Other Measures of Roughness and Riding Quality. In *Transportation Research Record 1084*, TRB, National Research Council, Washington, D.C., 1986.
6. K. W. Guenther. *Predictive Models for Vehicle Operating Consequences*. M.S. thesis. Civil Engineering Department, Massachusetts Institute of Technology, Cambridge, 1969.
7. F. Moavenzadeh et al. *Highway Design Standards Study Phase I: The Model*. Staff Working Paper 96. The World Bank, Washington, D.C., 1971.
8. A. D. St. John and D. R. Kobett. *NCHRP Report 185: Grade Effects on Traffic Flow Stability and Capacity*. TRB, National Research Council, Washington, D.C., 1978, 110 pp.
9. E. C. Sullivan. *The Slip Energy Approach to Vehicle-Road Interactions: Background and Prospects*. Working Paper 79-6. Institute of Transportation Studies, University of California, Berkeley, Jan. 1979.
10. A. Brodin, G. Gynnerstedt, and G. Levander. *A Program for Monte Carlo Simulation of Vehicle along Two-Lane Rural Roads; An Application of Structured Programming Technique and Simula-67 Language*. No. 143. National Road and Traffic Research Institute, Linkoping, Sweden, 1979.
11. L. E. Dickson. *New First Course in the Theory of Equations*. John Wiley, New York, 1957.
12. T. D. Gillespie. *Technical Consideration in the Worldwide Standardization and Road Roughness Measurement*. Report UM-HSRI-81-28. Highway Safety Research Institute, University of Michigan, Ann Arbor, 1981.
13. T. D. Gillespie, M. W. Sayers, and L. Segel. *NCHRP Report 228: Calibration of Response-Type and Roughness Measuring Systems*. TRB, National Research Council, Washington, D.C., 1980, 81 pp.
14. J. R. Benjamin and C. A. Cornel. *Probability, Statistics and Decision for Civil Engineers*, McGraw-Hill, New York 1970.
15. N. L. Johnson and S. Kotz. Continuous Univariate Distributions. Vol. I. Houghton Mifflin, Boston, Mass., 1970.
16. N. A. J. Hastings and J. B. Peacock. *Statistical Distributions: A Handbook for Students and Practitioners*. John Wiley, New York, 1974.
17. M. Ben-Akiva and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Predict Travel Demand*. MIT Press, Cambridge, Mass., 1985.
18. T. A. Domencich and D. McFadden. *Urban Travel Demand: A Behavioral Analysis*. North-Holland Publishing Co., Amsterdam, The Netherlands, 1975.
19. D. McFadden. Qualitative Response Models. In *Advances in Econometrics*, Chapter 1. W. Hildebrand (ed.), Cambridge University Press, Cambridge, England, 1982.
20. J. S. Williams. Efficient Analysis of Weibull Survival Data from Experiments on Heterogeneous Patient Populations. *Biometrica*, Vol. 34, 1978, pp. 209–222.
21. C. Gourieroux et al. Pseudo Maximum Likelihood Methods: Theory. *Econometrica*, Vol. 52, 1984, pp. 681–700.
22. E. Malinvaud. *Statistical Methods of Econometrics*. 3rd ed. North-Holland Publishing Co., Amsterdam, The Netherlands, 1980.
23. F. S. Koppelman. *Aggregation in Economic Analysis: An Introductory Survey*. Princeton University Press, Princeton, N.J., 1964.
24. D. McFadden. Econometric Models for Probabilistic Choice. In *Structural Analysis of Discrete Choice Data with Econometric Applications*. C. F. Manski and D. McFadden (eds.), MIT Press, Cambridge, Mass., 1981.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Comparative Analysis of Models for Estimating Delay for Oversaturated Conditions at Fixed-Time Traffic Signals

W. B. CRONJÉ

To optimize a fixed-time traffic signal, a model is required to estimate with sufficient accuracy the measure of effectiveness necessary for the optimization process. Suitable models have been developed for the degree of saturation (x) in the range $0 < x < 0.9$. Reliable models have also been developed for the zone of the degree of saturation $x > 1.1$. In this zone, traffic can be treated deterministically. However in the range $0.9 \leq x$ 1.10, the deterministic approach falls and a model should be based on the probabilistic approach to traffic flow. The ideal model should be applicable over the entire range of the degree of saturation. Only two such models have been encountered in the literature. Because of shortcomings of these models and the lack of a reliable model in the transition zone from undersaturation to oversaturation, an alternative model was developed by Cronjé (*Transportation Research Record 905*, 1983). This model is based on a Markov process and the geometric probability distribution, and is referred to in this paper as the M Geom Model. In this paper, the M Geom Model is compared with the models developed by Mayne and Catling on a cost basis. Monetary rates are assigned to the measures of effectiveness, namely, total delay and number of stops, for a wide range of cycle lengths, flows, and degree of saturation. The results indicate that the M Geom Model estimates cost more accurately and is consequently recommended for optimizing fixed-time traffic signals.

The predominant equations for estimating delay for undersaturated conditions have been developed by Webster, Miller, and Newell. However, these equations are only reliable for the degree of saturation (x) in the range $0 < x <$ about 0.9. However, because of oversaturated conditions existing for periods during the peak hours in urban conditions, there is a need for a model applicable over the entire range of the degree of saturation. Three such models have been developed by Mayne *(1)*, Catling *(2)*, and Cronjé *(3)*.

Catling adapted equations of classical queuing theory to oversaturated traffic conditions and developed comprehensive queue length and delay formulas that can represent the effects of junctions in a time-dependent model.

Mayne used some results on transient queuing theory to derive comprehensive queue and delay formulas for use in a procedure to optimize traffic signal settings during peak periods. He transformed existing formulas into formulas using the Poisson distribution, then applied appropriate statistical distribution approximations and other techniques of numerical analysis.

Cronjé treated traffic flow through a fixed-time signal as a Markov process and developed equations for estimating delay, queue length, and stops. The properties of the geometric probability distribution were then applied to these equations to obtain simpler equations, reducing computing time. The equations were then modified to further reduce computing time without sacrificing too much accuracy.

The measures of effectiveness used in this paper to compare the three models are delay and stops, but Catling and Mayne did not derive equations for estimating stops. However, the number of stops for the Mayne and Catling models can be obtained from their queue length equations by using the definition that the number of stops is the number of vehicles arriving while there is a queue plus the overflow of vehicles at the beginning of the cycle.

The probabilistic model used to describe the traffic arrivals at the intersection is the Poisson distribution because the Mayne and Catling models are both based on the Poisson distribution.

The basis of comparison of the models is simulation. Traffic arrivals at a signalized intersection is a stochastic process. If a traffic count is taken on a specific day for a particular period at an intersection and the delay and stops calculated, and if a count is then repeated for the same time period the next day, it is possible that the values obtained for delay and stops may differ appreciably. Consider the following set of arrivals generated that were obtained over 15 cycles for the following prescribed data:

- Cycle length: $c = 50$ sec;
- Degree of saturation: $x = 1.05$;
- Flow: $q = 500$ vehicles per hour; and
- Ratio of variance to mean of arrivals per cycle: $I = 0.5$.

Arrivals (a): 8, 8, 6, 6, 8, 8, 7, 8, 8, 5, 9, 8, 9, 8, 10 Σ 116. Initial queue length is 2 vehicles; average number of departures per cycle is 7.4 vehicles.

The equation for obtaining the overflow at the end of a cycle is as follows:

$$K = B + Z - V \tag{1}$$

where

K = overflow at the end of the cycle,
B = overflow at the beginning of the cycle,
Z = arrivals per cycle, and
V = departures per cycle.

Department of Civil Engineering, University of Stellenbosch, Stellenbosch 7600, Republic of South Africa.

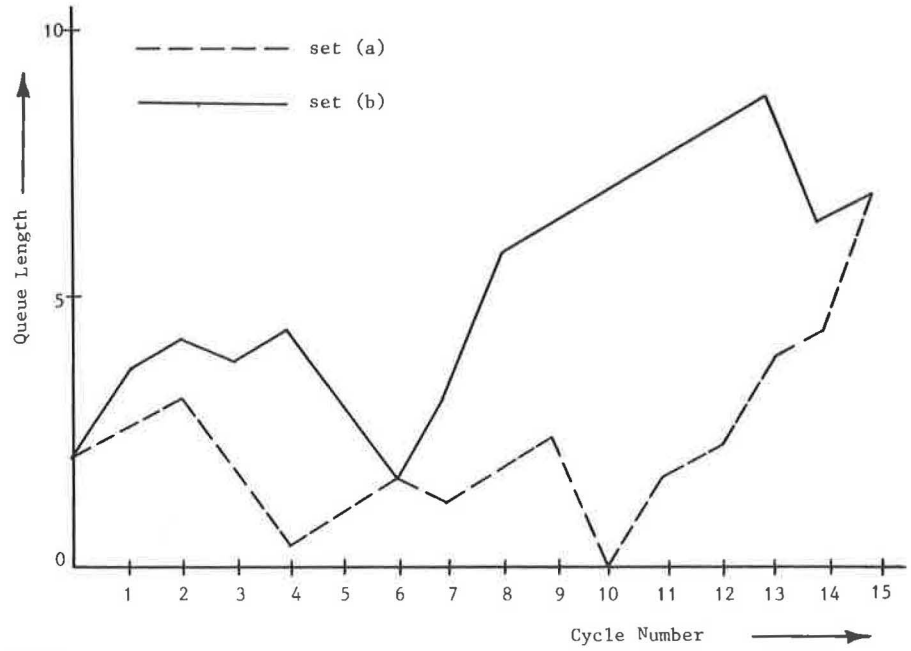


FIGURE 1 Queue length at end of 15 cycles.

Applying Equation 1 to the arrivals generated, the following values for K are obtained at the end of 15 cycles:

2.6, 3.2, 1.8, 0.4, 1.0, 1.6, 1.2, 1.8, 2.4, 0.0, 1.6, 2.2, 3.8, 4.4, 7.0.

Suppose the order of arrivals is changed to the following:

Arrivals (b): 9, 8, 7, 8, 6, 6, 9, 10, 8, 8, 8, 8, 8, 5, 8 Σ = 116.

The following values for K are now obtained:

3.6, 4.2, 3.8, 4.4, 3.0, 1.6, 3.2, 5.8, 6.4, 7.0, 7.6, 8.2, 8.8, 6.4, 7.0.

A plot of the two sets of overflow is shown in Figure 1. Although the area under the curves is not the delay, it is directly related to delay and obviously the delay differs appreciably for the two sets.

Calculating delay as the area under the queue length curve and calculating stops from the definition previously defined, the values for the two sets of arrivals obtained are given in Table 1 alongside those obtained by the M Geom Model.

From the values indicated in Table 1 it is therefore clear that the order of the arrivals affects delay and stops considerably. If actual counts have to be used, a fairly large sample will be required to properly estimate the population measures of effectiveness. On the other hand, the larger the sample, the longer is the calendar period over which the counts are taken; during this period, changes in the traffic flow conditions may occur, biasing the sample. This problem is eliminated with computer simulation, which shall therefore be applied.

Traffic arrivals are generated by random numbers over a wide range of cycle lengths, flows, and degrees of saturation. To introduce as large a degree of variation as possible into the sample, arrivals are generated for the binomial and negative binomial distributions for values of I varying from 0.5 to 1.5, excluding the case of I = 1, which gives the Poisson distribution. For each model the cost, obtained from assigning monetary rates to the measures of effectiveness—namely, delay and stops—is compared with the actual cost obtained for the arrivals generated. The results are analyzed and it is found that in the case of the M Geom and Catling models the cost difference follows the normal distribution at the 5 percent level. The models are compared extensively and the M Geom Model is found to estimate cost more accurately than do the Mayne and Catling models.

TABLE 1 DELAY AND STOPS FOR TWO SETS OF GENERATED DATA

Delay			Stops			
Set	Actual	M Geom	Actual	M Geom	ΔD (%)	ΔN (%)
a	3,770	6,797	150	210	80.3	40.0
b	6,053	6,797	192	210	12.3	9.4

Note: ΔD = difference in delay and ΔN = difference in stops.

ANALYSIS

The equations used in the Mayne, Catling, and M Geom models for estimating delay, stops, and overflow shall now be given.

Mayne Model

Mayne used some results on transient queuing theory to derive comprehensive queue and delay formulas for use in a pro-

cedure to optimize traffic signal settings during peak periods. He transformed existing formulas into formulas using the Poisson distribution, and then applied appropriate statistical distribution approximations and other techniques of numerical analysis.

Queue Length Formulas

After applying further approximations, Mayne obtained the following queue length formulas.

Rising Queue The rising occurs under the following conditions:

$x \geq 1$; $x < 1$, and $Q < Q_e$.

$$Q(t) = (w^2 \cdot u + k_1)\,(w^2 \cdot u + k_2)^{-1/2}\,u^{1/2} - w \cdot u \quad (2)$$

$$Q_e \doteq x/[2(1-x)] = 1/4 \cdot w \quad (3)$$

where

x = degree of saturation,
Q = queue length at start of period under consideration,
Q_e = equilibrium queue length,
$Q(t)$ = queue length at time t,
w = $(1-x)/(2 \cdot x)$,
u = $q \cdot t$,
k_1 $\doteq$ 0.3417, and
k_2 $\doteq$ 0.1834.

Falling Queue The falling queue occurs under the following conditions:

$x < 1$ and $Q > Q_e$

$$Q(t) = (w^2 \cdot u^2 + 0.5 \cdot u + k_5 \cdot w^{-2})^{1/2} - w \cdot u \quad (4)$$

where k_5 is 0.1202.

For rising queues, if $Q \neq 0$, then the origin, where the queue length is zero, has to be determined. This condition is shown in Figure 2. In the case of the falling queue, the zero origin is determined similarly except that the queue length at the origin is larger than Q.

Formulas for Obtaining Zero Origin

The formulas for obtaining zero origin are as follows.

Rising Queue The rising queue occurs under the following conditions:

$0 < x < 1$ and $Q < Q_e$, or $x > 1$.

$$R(Q) = (C - B)/(2 \cdot A \cdot q) \quad (5)$$

where

$R(Q)$ = inverse function of $Q(t)$, which gives the approximate value of the time parameter for which the mean queue has value Q;
C = $(B^2 + 4 \cdot k_2 \cdot A \cdot Q^2)^{1/2}$;
B = $k_1^2 - 2k_2 \cdot w \cdot Q - w^2Q^2$;
A = $w^2(0.5 - 2 \cdot w \cdot Q)$; and
q = average arrival rate of traffic.

Falling Queue The falling queue occurs under the following conditions:

$0 < x < 1$ and $Q > Q_e$.

$$R(Q) = (k_5/w^2 - Q^2)/q(2 \cdot w \cdot Q - 0.5) \quad (6)$$

Having determined $R(Q)$ from Equations 5 and 6, $Q(t)$ is then obtained from Equations 2 and 4 such that

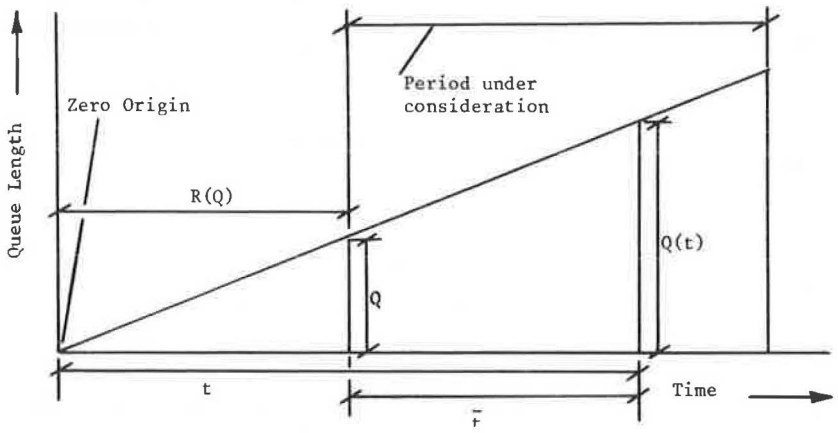


FIGURE 2 Determination of zero time origin for a rising queue.

$$Q(t) = Q[R(Q) + \bar{t}\,]$$

where $\bar{t}$ is the time from the start of the period under consideration to where the queue length is required.

Delay Formulas

The delay formulas are as follows:

Oversaturation For oversaturation, the following conditions are implied:

$x > 1$, $x < 1$, and $Q > Q_e$.

Delay during the peak period $[T, T + L]$, that is, of duration L, is given by

$$R(T,T + L) = S(T,T + L) + 0.5 \cdot c\,(1 - \lambda)$$
$$[W + Q(T + L) - Q(T)] \qquad (7)$$

and

$$S(T,T + L) = \int_T^{T+L} Q(t) \cdot dt = c\,[0.5(Q_o + Q_n) + Q_1 + Q_2 + \ldots + Q_{n-1}]$$

where

- W = $q \cdot L$ = total approach flow during period under consideration,
- L = length of period under consideration,
- Q_o = $Q(T)$ = queue length at start of period under consideration,
- Q_n = $Q(T + L)$ = queue length at end of period under consideration,
- c = cycle length,
- λ = g/c = proportion of cycle that is effectively green, and
- g = effective green time.

Undersaturation For undersaturation, the following conditions are implied:

$x < 1$ and $Q < Q_e$.

Mayne, however, suggests the use of Webster's approximate Equation 8 for average delay for $0 < x <$ approximately 0.9 and $Q < Q_e$.

$$d = 0.9\,[c(1 - \lambda)^2/2(1 - \lambda \cdot x) + x^2/2 \cdot q(1 - x)] \qquad (8)$$

where d is average delay per cycle.

It follows that

$$D = d \cdot q \cdot c$$

where D is the total delay per cycle.

Catling Model

Catling adapted existing equations of classical queuing theory to oversaturated traffic conditions and developed comprehensive queue and delay formulas that can represent the effects of junctions in a time-dependent model.

Queue Length Equations

The queue length equations are as follows.

Rising Queue The same conditions for a rising queue apply as for the Mayne model. As in the case of the Mayne model, if there is a queue at the start of the period under consideration, the origin at which the queue length is zero has to be determined. For the Mayne model the origin is calculated directly but in the Catling model it has to be obtained by trial and error.

$$G(x, t) = [(\beta^2 + 2 \cdot x^2 \cdot Q^2 \cdot t^2 \cdot \alpha \cdot C)^{1/2} - \beta]/\alpha \qquad (9)$$

where

- $G(x, t)$ = queue length at time t with degree of saturation x,
- t = time from zero origin to point where queue length is required,
- α = $2(Q \cdot t - C)$,
- β = $Q \cdot t\,[Q \cdot t\,(1 - x) + 2 \cdot C \cdot x]$,
- Q = $s \cdot g/c$, and
- C = 0.55.

The equilibrium queue length, $Gl(x)$, with degree of saturation x, derived by Catling, is given by

$$Gl(x) = C \cdot x^2/(1 - x) \qquad (10)$$

Falling Queue The same conditions for a falling queue apply as for the Mayne Model. In this case, a zero origin is not determined. The queue length formula is as follows:

$$G(x,t,G_o) = G_o - Q \cdot t(1 - x) \qquad (11)$$

where t is the time measured from the start of the period under consideration, and G_o is the queue length at the start of the period under consideration.

Delay Formulas

Catling gives average delay formulas as follows.

Rising Queue The following conditions are implied: $x \geq 1$, $x < 1$, and $G_o \leq Gl(x)$.

$$d(x,t,G_o) = [(t + t') \cdot D(x,t + t') - t' \cdot D(x,t')]/t \quad (12)$$

where

- t = length of time period under consideration;
- t' = time from the zero origin to start of time period under consideration;
- $d(x,t,G_o)$ = average delay over period under consideration with degree of saturation x and initial queue length G_o;
- $D(x,t + t')$ = average delay over period $(t + t')$, that is, from the zero origin to the end of the period under consideration; and
- $D(x,t')$ = average delay over period t', that is, from the zero origin to the start of the period under consideration.

In general,

$$D(x,t) = A + (N - M)/4 \cdot Q \quad (13)$$

where

- $M = 2 \cdot C + t \cdot Q(1 - x)$,
- $N = (M^2 + 8 \cdot C \cdot Q \cdot t \cdot x)^{1/2}$,
- $D(x,t)$ = average delay over time period t with degree of saturation x, and
- t = time period from zero origin.

Falling Queue The following conditions are implied: $x < 1$ and $G_o > Gl(x)$.

In addition, the following conditions are treated:

$$t \leq [G_o - Gl(x)]/Q(1 - x) \text{ and } t > [G_o - Gl(x)]/Q(1 - x)$$

$$t \leq [G_o - Gl(x)]/Q(1 - x)$$

The average delay equation is as follows:

$$d(x,t,G_o) = A - 0.5 \cdot t(1 - x) + G_o/Q \quad (14)$$

where A is $0.5 \cdot c\ (1 - \lambda)^2$ and t is length of time period under consideration.

$$t < [G_o - Gl(x)]/Q(1 - x)$$

$$d(x,t,G_o) = A + Gl(x)/x \cdot Q + [G_o - Gl(x)]\{x[G_o + Gl(x)] - 2 \cdot Gl(x)\}/2 \cdot x \cdot Q^2 \cdot t(1 - x) \quad (15)$$

For use in a comparative analysis, Equations 8, 12, 14, and 15 have to be transformed to total delay such that

$$D(x,t,G_o) \doteq d(x,t,G_o) \cdot q \cdot t$$

where $D(x,t,G_o)$ is the total delay over time period t under consideration with degree of saturation x and initial queue length G_o, and q is the average arrival rate of traffic over time period t.

M Geom Model

The equations of the M Geom Model are as follows:

$$E(K) = E(B) + E(Z) - E(V) - \sum_v p(v) \sum_{z=o}^{v-1} p_Z(z)[E(B)(1 - F^{v-z}) - v + z] \quad (16)$$

$$E(N) = E(B) + E(Z) + \sum_v p_V(v) \sum_{z=o}^{v-1} p_Z(z)\ (z/c) + (v/g - z/c)[E(B)(1 - F^{v-z}) + z - v] \quad (17)$$

$$E(D) = E(B)c + [E(Z)c - E(V)g]/2 + \sum_v p_V(v) \sum_{z=o}^{v-1} p_Z(z)\{E(B)[2 - F^{v-z}(1 + F)] + (v - z)^2 - (v - z + 1)^2 F\}/2(1 - F)(v/g - z/c) \quad (18)$$

where

- $E(K)$ = expected value of the overflow at the end of the cycle,
- $E(N)$ = expected value of the number of stops per cycle,
- $E(D)$ = expected value of the total delay per cycle,
- $E(B)$ = expected value of the overflow at the beginning of the cycle,
- $E(Z)$ = expected value of the arrivals per cycle,
- $E(V)$ = expected value of the maximum number of departures per cycle, and
- $F = E(B)/[1 + E(B)]$.

The range of the degree of saturation is divided into three zones as follows:

- Zone 1: $0 < x < 0.9$,
- Zone 2: $0.9 \leq x < 1.009$, and
- Zone 3: $x \geq 1.009$.

In some zones, Equations 16, 17, and 18 shall be applied in a modified form.

Zone 1: $0 < x < 0.9$

In this zone there are two possibilities, namely, $B < Q_o$ and $B > Q_o$, where B is this queue length at the beginning of the period under consideration, and Q_o is the equilibrium overflow, that is, the stabilizing value of the queue length after some time.

Cronjé found that the following Newell equations are the most accurate for estimating delay and stops for undersaturated conditions *(4)*:

$$d = c(1 - \lambda)^2/2(1 - \lambda \cdot x) + Q_o/q$$

$$D = d \cdot q \cdot c \quad (19)$$

$$N = q[(q \cdot r + Q_o)/(s - q) + r] + Q_o \quad (20)$$

where

d = average delay (sec per vehicle);
D = total delay per cycle (vehicle seconds);
N = number of stops per cycle;
Q_o = $I \cdot H(\mu)x/2\ (1 - x)$;
c = cycle length (sec);
g = effective green time (sec);
λ = g/c = proportion of the cycle that is effectively green;
I = ratio of the variance to the average of the arrivals per cycle;
x = degree of saturation, the ratio of the average number of arrivals per cycle to the maximum number of departures per cycle;
q = average arrival rate (vehicles per second);
s = saturation flow (vehicles per second);
$H(\mu)$ = $e^{-(\mu+\mu^2/2)}$;
μ = $(1 - x)(s \cdot g)^{1/2}$; and
r = effective red time (sec).

The method of obtaining delay and stops in Zone 1 shall now be given.

For $B < Q_o$: When there is an increase in flow, delay does not increase instantaneously to the new value but fluctuates until it stabilizes at the end of the nonstationary zone, as indicated in Figure 3.

Determine delay and stops from Equations 19 and 20. Now also determine delay and stops by using Equations 16–18 for the initial number of cycles in the nonstationary zone. Check the delay obtained for each cycle against 0.9 (delay obtained by using Equation 19). As soon as the calculated delay exceeds this value, the calculation of delay and stops ceases and the sums of delay and stops for the period under consideration are as follows:

Σ Delay = Σ delay over the first number of cycles + $M \cdot$ (delay obtained by using Equation 19), and

Σ Stops = Stops over first number of cycles + $M \cdot$ (stops obtained by using Equation 20).

where

M = $N - Y$,
N = number of cycles in this period under consideration, and
Y = number of cycles until calculated delay: ≥ 0.9 (delay obtained by using Equation 19).

For $B > Q_o$: This case is treated exactly as when $B < Q_o$ except that after the delay for each cycle has been obtained, it is checked against the following value:

1.1 · (delay obtained by using Equation 19)

and Y is the number of cycles until the calculated delay is $\leq$ 1.1 · (delay obtained by Equation 19).

Zone 2: $0.9 \leq x < 1.0009$

Also in this case, delay and stops are determined from Equation 16–18 for the period under consideration.

Zone 3: $x \geq 1.0009$

In this zone, Equation 16 is modified to the following:

$$E(K) = E(B) + E(Z) - E(V) + H/E(B) \quad (21)$$

where

$$H = E(B)_1 \sum_v p_V(v) \sum_{z=o}^{v-1} p_Z^{(z)}[E(B)_1 (1 - F^{v-z}) + z - v]$$

$E(B)_1$ = expected queue length at the beginning of the first cycle.

In this zone, stops and delay are evaluated from Equations 17 and 18 but the summation terms are omitted.

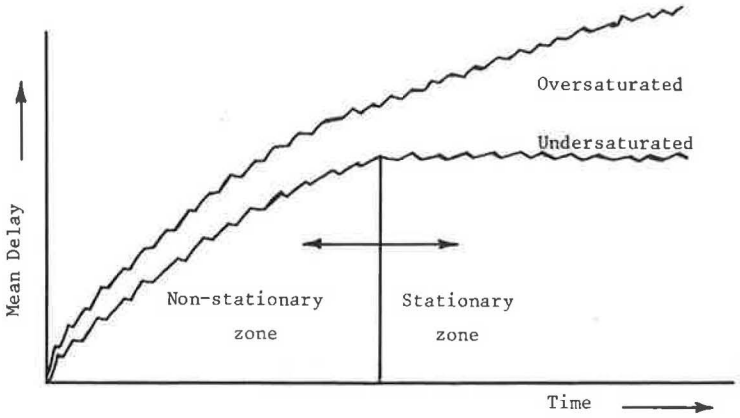


FIGURE 3 Transition from nonstationary to stationary zone as flow increases.

TABLE 2 DATA FOR GENERATING ARRIVALS

Input				Output			
c	x	q	I	c	x	q	g
50	1.05	500	0.5	50	1.05	542.4	14.4
			0.6			547.2	14.5
			0.7			556.8	14.7
			0.8			499.2	13.2
			0.9			552.0	14.6
			1.1			499.2	13.2
			1.2			537.6	14.2
			1.3			432.0	11.4
			1.4			480.0	12.7
			1.5			465.6	12.3
50	1.10	650	0.5	50	1.10	633.6	16.0
			(and so forth)				

Note: c = cycle length; x = degree of saturation; q = average arrival rate; I = ratio of variance to mean of arrivals per cycle; and g = effective green time (sec).

TABLE 3 FITTING DATA WITH NORMAL DISTRIBUTION FOR M GEOM

Group	Observed Frequency (O)		Theoretical Frequency (T)		$(O-T)^2/T$
< −40	6	12	1.92	10.30	0.28
−40 to −30	6		8.38		
−30 to −20	23		27.14		0.63
−20 to −10	54		57.95		0.27
−10 to 0	81		78.62		0.07
0 to 10	80		76.32		0.18
10 to 20	51		45.70		0.61
20 to 30	17		17.95		0.05
30+	2		6.02		2.68
Σ =	320		320		4.77

Note: $\bar{x} = -1.906$ and $\bar{s} = 15.173$.

DATA GENERATED

The data generated are based on the following values for c, I, x, and q:

c (sec) = 50, 60, 70, 80, 90, 100, 110, 120
I = 0.5, 0.6, 0.7, 0.8, 0.9, 1.1, 1.2, 1.3, 1.4, 1.5

The degree of saturation and the flow are assigned the following paired values:

x	q (veh/hr)
1.05	500
1.10	650
1.15	800
1.20	900

Arrivals are generated over 15 cycles. The arrivals are converted to flow. The cycle length and degree of saturation are kept the same and the corresponding effective green time is calculated. To limit the length of the paper, only input and output data obtained for c = 50 sec are given in Table 2. A total of 320 sets of data are obtained.

The equations used in the M Geom, Mayne, and Catling models are applied to the data generated. The cost difference is then checked against the normal distribution.

M Geom

The histogram of a sample from arrivals generated for the M Geom Model is shown in Figure 4 and the processed data are given in Table 3.

The number of groups = n = 8. The number of degrees of freedom lost = 3. Therefore, the number of degrees of freedom $f = n - 3 = 8 - 3 = 5$.

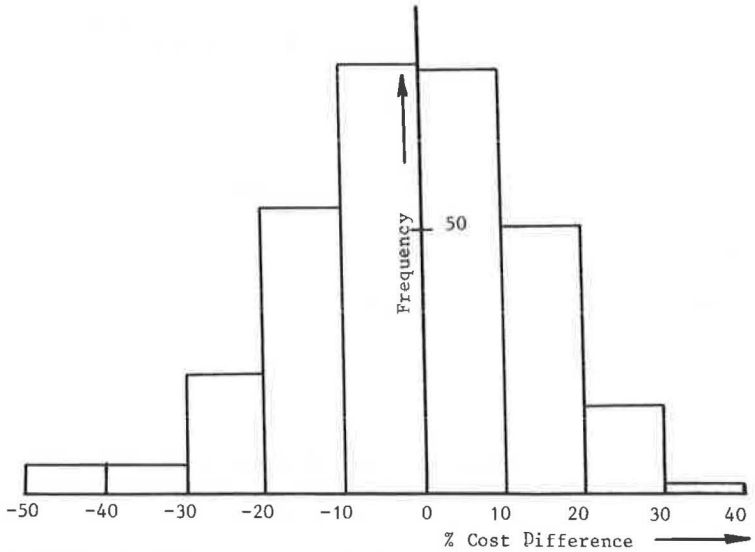


FIGURE 4 Histogram of sample from arrivals generated for M Geom Model.

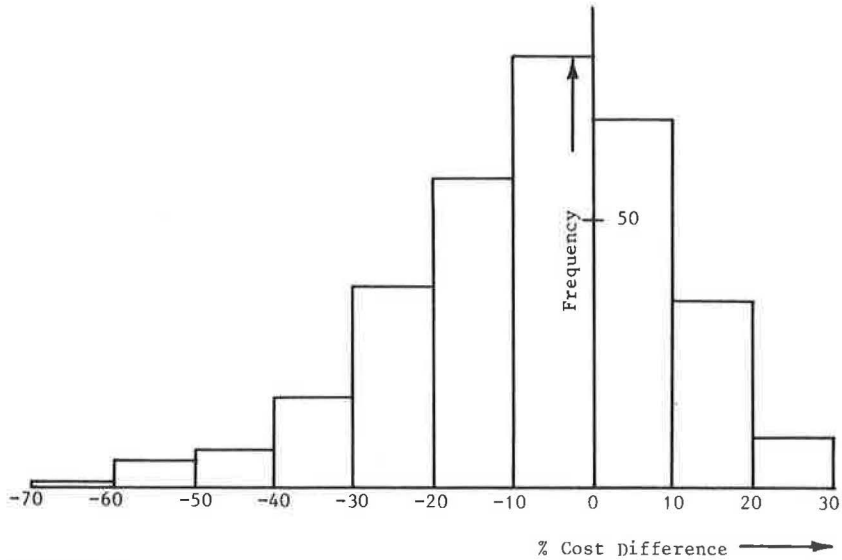


FIGURE 5 Histogram of sample from arrivals generated for Mayne Model.

At the 5 percent level $\chi^2_{5;0.05} = 11.07$

$\Sigma(0 - T)^2/T = 4.77 < 11.07$

Therefore, there is insufficient evidence to reject the hypothesis that the percentage cost difference follows the normal distribution. The 95 percent confidence interval is given by

$x \pm 1.96 \cdot \bar{s}$

Therefore, the confidence interval is

$[-1.906 \pm 1.96 \cdot 15.173]$

$= [-31.64\%; 27.83\%]$

The following monetary rates are used: $3.1 \cdot 10^{-2}$ rand per stop; $1.74 \cdot 10^{-4}$ rand per vehicle second for total delay. The rand is the unit of currency in the Republic of South Africa [1 rand = \$0.62 (1984)]. These rates are from research to be published by the National Institute for Road and Transport Research, Republic of South Africa.

Mayne Model

The histogram of a sample from arrivals generated for the Mayne Model is shown in Figure 5 and the processed data are given in Table 4.

The number of groups $n = 8$. The number of degrees of freedom lost = 3. Therefore, the number of degrees of freedom = 8 – 3 = 5.

At the 5 percent level $\chi^2_{5;0.05} = 11.07$.

$\Sigma(0 - T)^2/T = 12.07 > 11.07$

Therefore, there is insufficient evidence to accept the hypothesis that the percentage cost difference follows the normal distribution.

TABLE 4 FITTING DATA WITH NORMAL DISTRIBUTION FOR MAYNE MODEL

Group	Observed Frequency (O)		Theoretical Frequency (T)		$(O - T)^2/T$
< –60	1		0.29		
–60 to –50	5	13	1.60	8.58	2.28
–50 to –40	7		6.69		
–40 to –30	17		20.26		0.52
–30 to –20	38		44.64		0.99
–20 to –10	58		67.46		1.33
–10 to 0	81		73.47		0.77
0 to 10	69		57.12		2.47
10 to 20	35		31.97		0.29
20+	9		16.51		3.42
Σ =	320		320.01		12.07

Note: $\bar{x} = -7.438$ and $\bar{s} = 16.862$.

Because the Mayne Model does not follow the normal distribution at the 5 percent level, the confidence interval shall be obtained from the histogram in Figure 5 by locating the value of the variant at each tail for 2.5 percent of the area. The number of observations given by 2.5 percent is given by

$2.5 \cdot 320/100 = 8$

The 95 percent confidence interval is given by

$[-47.5\%; 20.0\%]$

Catling Model

The histogram of a sample from arrivals generated for the Catling Model is shown in Figure 6, and the processed data are given in Table 5.

The number of groups = $n = 8$. The number of degrees of

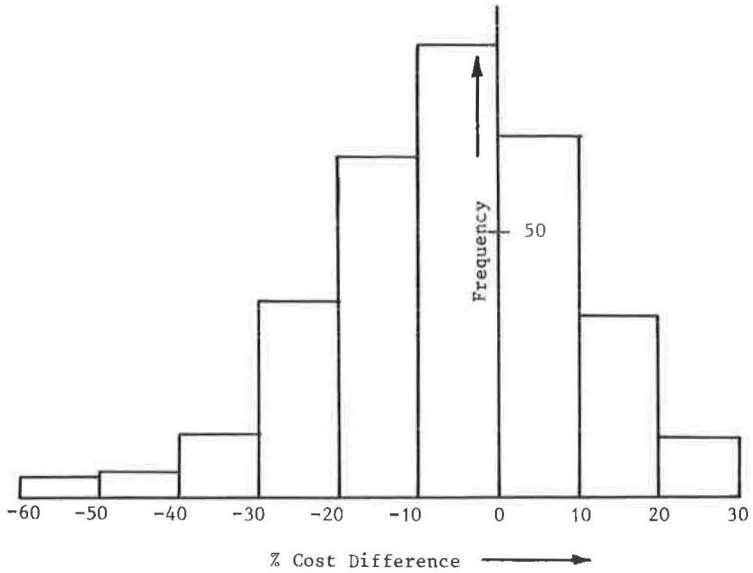


FIGURE 6 Histogram of sample from arrivals generated for Catling Model.

freedom lost = 3. Therefore, the number of degrees of freedom = 8 – 3 = 5.

At the 5 percent level $\chi^2_{5;0.05} = 11.07$.

$\Sigma(0 - T)^2/T = 6.59 < 11.07$

Therefore, there is insufficient evidence to reject the hypothesis that the percentage cost difference follows the normal distribution.

The 95 percent confidence interval is given by

$\bar{x} \pm 1.96 \cdot \bar{s}$

Therefore, the confidence interval is

$[-6.406 \pm 1.96 \cdot 15.803]$

$= -37.4\%; 24.6\%$

TABLE 5 FITTING DATA WITH NORMAL DISTRIBUTION FOR CATLING MODEL

Group	Observed Frequency (O)		Theoretical Frequency (T)		$(O - T)^2/T$
< – 50	4	9	0.93	5.31	
–50 to –40	5		4.38		2.56
–40 to –30	12		16.48		1.22
–30 to –20	37		40.58		0.32
–20 to –10	64		68.51		0.30
–10 to 0	85		80.03		0.31
0 to 10	68		61.34		0.65
10 to 20	34		32.54		0.07
20+	11		15.20		1.16
Σ =	320		319.99		6.59

Note: $\bar{x} = -6.406$ and $\bar{s} = 15.803$.

The confidence intervals for the three models have been developed for the 320 sets of generated data. However, it is possible that on the average one model may appear preferable but that for specific ranges of the data one of the other models may estimate cost more accurately. To investigate this possibility, the variation in percentage cost difference is examined against the following:

1. Variation in cycle length (c),
2. Variation in paired sets of flow (Q) and degree of saturation (x), and
3. Variation in the variance to mean ratio of the arrivals per cycle (I).

The values obtained are given in Tables 6–8 and plotted in Figures 7–9. Because the negative values for percentage cost difference actually indicate an overestimation of cost, these values shall be plotted on the positive axis in Figures 7–9.

First, the variation of percentage cost difference against cycle length variation is examined for the M Geom, Mayne, and Catling models. The values obtained are given in Table 6.

TABLE 6 PERCENTAGE COST DIFFERENCE AGAINST CYCLE LENGTH VARIATION FOR THE THREE MODELS

Cycle Length (c)	M Geom	Mayne	Catling
50	–0.72	–6.27	–5.02
60	–6.93	–12.96	–11.78
70	–6.42	–12.57	–11.47
80	0.82	–4.93	–4.02
90	–1.17	–7.15	–6.19
100	1.30	–4.47	–3.67
110	1.19	–4.54	–3.78
120	–2.04	–8.12	–7.23

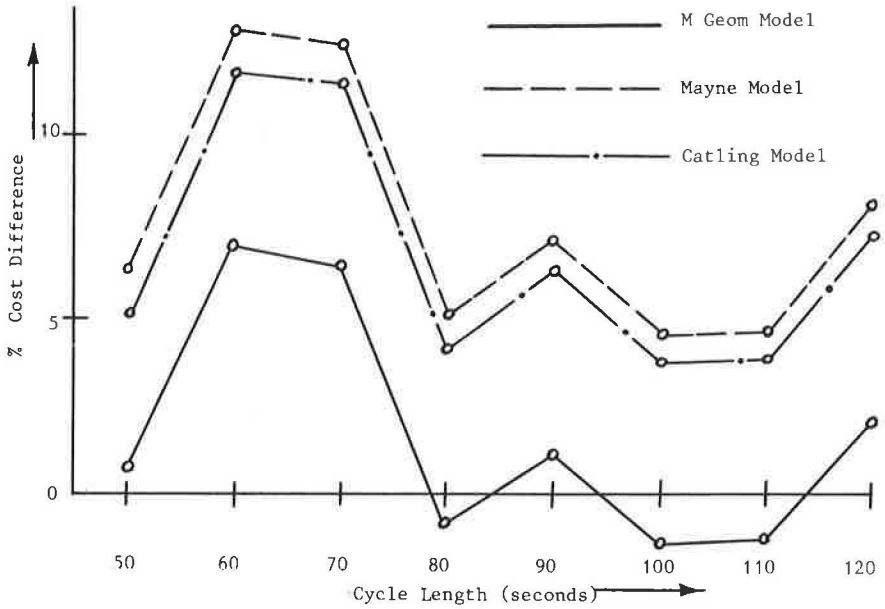


FIGURE 7 Percentage cost difference against cycle length variation for the three models.

TABLE 7 PERCENTAGE COST DIFFERENCE AGAINST FLOW AND DEGREE OF SATURATION FOR THE THREE MODELS

Flow (Q)	Degree of Saturation (x)	M Geom	Mayne	Catling
500	1.05	–6.64	–16.94	–12.63
650	1.10	1.34	–4.42	–2.89
800	1.15	–0.35	–4.46	–4.65
950	1.20	–1.34	–4.68	–6.41

The percentage cost difference given in Table 6 for the three models is plotted in Figure 7.

Second, the variation of percentage cost difference against paired values of flow and degree of saturation is examined for the M Geom, Mayne, and Catling models. The values obtained are given in Table 7. The percentage cost difference indicated in Table 7 for the three models is plotted in Figure 8.

Last, the variation of percentage cost difference against the ratio of the variance to the mean of the arrivals per cycle is examined for the M Geom, Mayne, and Catling models. The values obtained are given in Table 8. The percentage cost difference given in Table 8 for the three models is plotted in Figure 9.

DISCUSSION OF M GEOM, MAYNE, AND CATLING MODELS

The 95 percent confidence intervals and means for the three models for the data generated are as follows:

- M Geom Model: [–31.64; 27.83]; $\bar{x}$ = –1.906.
- Mayne Model: [–47.5; 20.0]; $\bar{x}$ = –7.438.
- Catling Model: [–37.4; 24.6]; $\bar{x}$ = –6.406.

Although the range in the confidence interval is approximately the same, the M Geom Model estimates cost on the average considerably closer than do the Mayne and Catling models. From an inspection of Figures 7–9 it is also clear that throughout the range of varying cycle length, flow and degree of saturation, and variance to mean ratio of arrivals per cycle, the M Geom Model estimates cost more accurately. The only exception is when $I > 1.45$, which is negligible. Computing time for the three models for the generation of the arrivals and for the computation of cost is as follows:

M Geom Model: 1 min 54.5 sec
Mayne Model: 1 min 52.8 sec
Catling Model: 1 min 53.8 sec

It is clear that in computing time for the three models, there is so little difference that there is no reason to prefer one model in particular. It is therefore concluded that the M Geom Model

TABLE 8 PERCENTAGE COST DIFFERENCE AGAINST I FOR THE THREE MODELS

I	M Geom	Mayne	Catling
0.5	–8.10	–14.46	–13.32
0.6	0.75	–5.02	–4.05
0.7	–6.27	–12.48	–11.36
0.8	–3.40	–9.32	–8.32
0.9	–4.82	–10.88	–9.87
1.1	–0.62	–6.44	–5.47
1.2	–0.20	–5.88	–5.05
1.3	0.41	–5.34	–4.38
1.4	0.52	–5.27	–4.29
1.5	4.28	–1.21	–0.32

Note: I = ratio of variance to mean of arrivals per cycle.

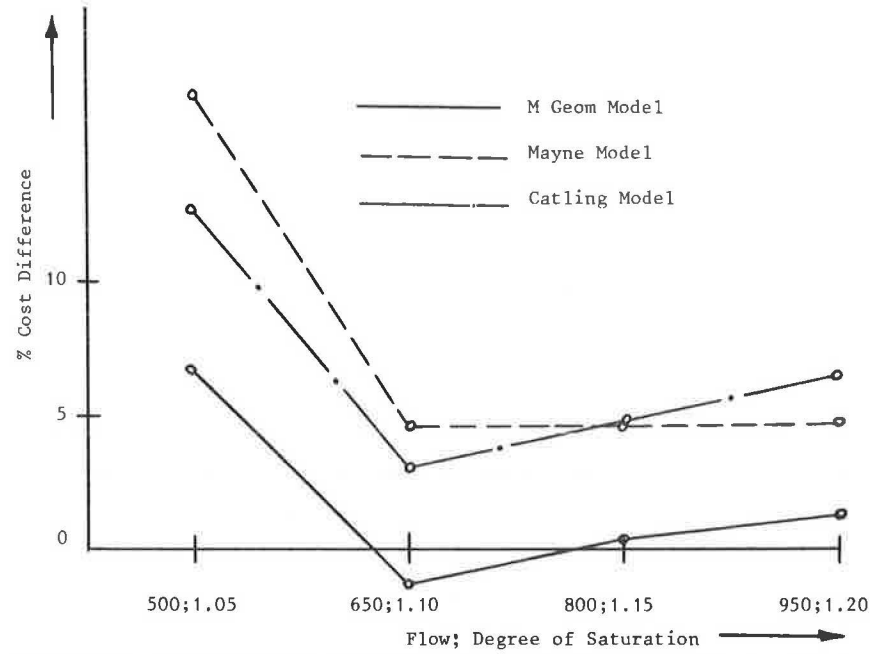


FIGURE 8 Percentage cost difference against flow and degree of saturation for the three models.

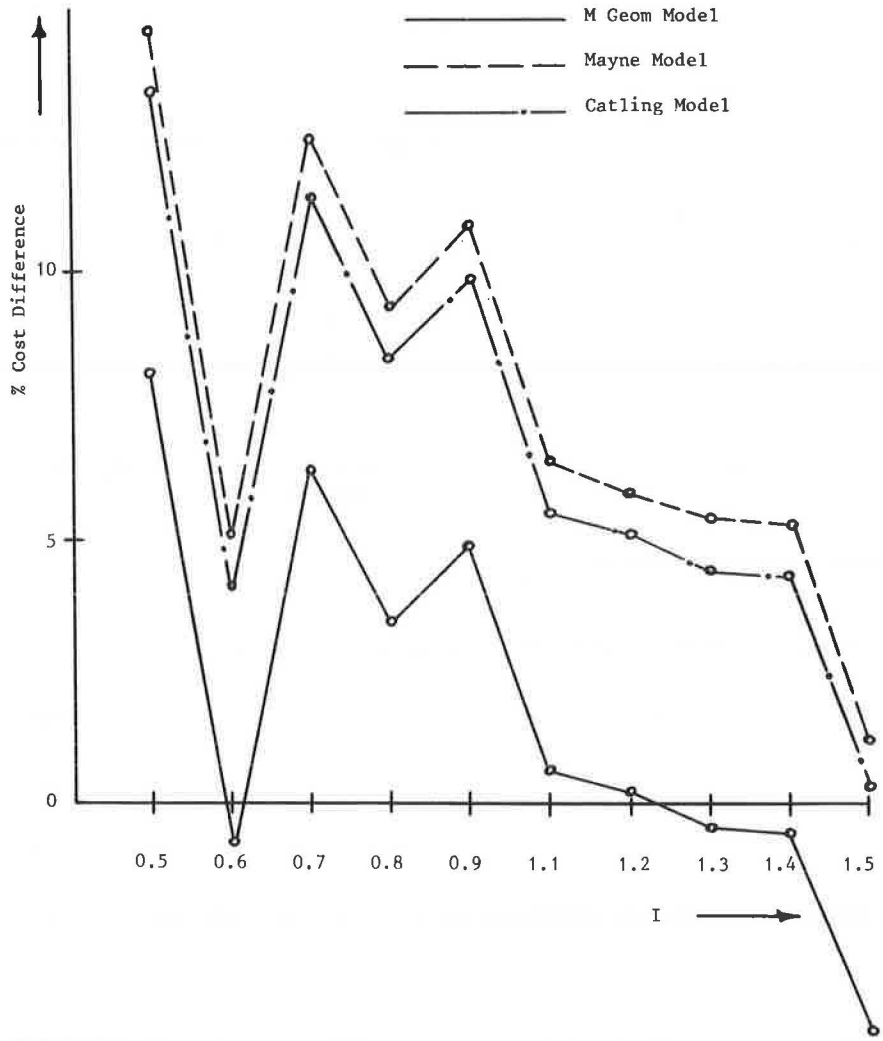


FIGURE 9 Percentage cost difference against variation in *I* for the three models.

is the preferred model for estimating cost in the optimization procedure for fixed-time traffic signalized intersections for undersaturated and saturated conditions.

REFERENCES

1. A. J. Mayne. "Traffic Signal Settings for Congested Traffic." *Proc., Engineering Foundation Conference on Research Directions in Computer Control of Urban Traffic Systems*, Calif., Feb. 1979.
2. I. Catling. "A Time-Dependent Approach to Junction Delays." *Traffic Engineering and Control*, Nov. 1977.
3. W. B. Cronjé. "Optimization Model for Isolated Signalized Traffic Intersections." In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp. 80–83.
4. W. B. Cronjé. "Analysis of Existing Formulas for Delay, Overflow, and Stops." In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp. 89–93.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Peaking Characteristics of Rural Road Traffic

PIETER W. JORDAAN AND CHRISTO J. BESTER

A methodology for estimating the traffic volume on rural roads for any of the highest thousand hours of the year is presented. The methodology requires estimates of the annual average daily traffic and a peaking characteristic. Also presented is the derivation of the latter from the data obtained from permanent traffic counters. It is found that the peaking characteristic is related to the fraction obtained by dividing the thirtieth highest hourly volume by the annual average daily traffic as well as by the average length of through trips on the road link to which it is applicable. It is hoped that this methodology will create a sound base for further research into traffic volume variations and design hour volume.

Peaking of rural traffic is a well-known phenomenon. How to take such peaking into account is less well established. Two approaches are in use in highway engineering practice, namely, the design hour approach and the flow regime approach.

The design hour approach was established in the United States by Peabody and Norman in 1941 *(1)*. They found that if the hourly traffic counts from a year's records were sorted from high to low and plotted against rank number, there was a "knee" at about the thirtieth highest hour. Since 1941 this concept of using some hourly volume to represent the peaking characteristic of the traffic on a road has received much support. It is embodied in many design textbooks *(2–4)*, usually of U.S origin.

In South Africa this concept is also well-known and applied by most road authorities *(5)*. Some local research has been done into the relationship between K [the fraction obtained by dividing the thirtieth highest hourly volume by the annual average daily traffic (AADT)] and the AADT *(6)*. To date, no real causal parameter for estimating K has been found. K is generally believed to vary as follows:

- Roads with a high proportion of recreational traffic: $K > 0.25$,
- Average roads: $0.15 < K \leq 0.25$, and
- Roads with little peaking: $K \leq 0.15$.

It has been observed that as AADT increases, there is a tendency for K to decline *(6)*.

The flow regime approach was established by Dawson in the United Kingdom in the late 1960s *(7)*. In this approach, the 8,760 hr of the year are divided into groups with constant flow characteristics. Dawson's original flow regimes for rural roads are given in Table 1. Over the years, this approach has been refined by the Transport and Road Research Laboratory in the United Kingdom *(8)*. It is also used in the COBA procedures for project evaluation *(9)*.

In South Africa, this approach is also used in the RODES2 program developed by Bester *(10)*, in which the flow regimes are differentiated by K, the fraction of the thirtieth highest hourly volume.

From the foregoing it is clear that a knowledge of hourly vehicular flow is important to the roads engineer. Not only is a knowledge of the flow during the thirtieth highest hour required, but the use of sophisticated project evaluation pro-

P. W. Jordaan, University of Pretoria, Lynwood Rd., Pretoria, Republic of South Africa, 0002. C. J. Bester, National Institute for Transport and Road Research, CSIR, Pretoria, Republic of South Africa.

TABLE 1 FLOW REGIMES FOR RURAL ROADS ACCORDING TO DAWSON

Flow Group	No. of Hours Represented (1)	Avg Flow in Veh/Hr as a Fraction of AADT (2)
1	3,800	0.010
2	3,420	0.054
3	1,160	0.083
4	380	0.115

Note: (1) × (2) summed over the four groups = 365 = number of days in the year.

cedures requires estimates of the vehicular flow in other ranked hours of the year.

Therefore, the purpose of this paper is to develop a methodology for the determination of hourly volumes on rural roads that is able to replace the current methodologies and that will create a sound base for further research into traffic volume variations.

RELATIONSHIP BETWEEN HOURLY VOLUME AND RANK NUMBER, N

Analysis and Development of Theory

If hourly counts at continuous counting stations are sorted from high to low and plotted versus rank number on a log–log scale, the plot between the tenth highest hour and the 1,000th highest hour forms a virtual straight line. This implies that for this region, the following relationship holds:

$$F = aN^b$$

where

F = hourly count/AADT,
a and b = calibration values, and
N = rank number.

Lines were fitted by means of least-squares techniques to determine a and b for the data from each of 65 different counting stations. Figure 1 shows an example of the hourly counts plotted at a continuous station that was counted for 18 hr/day. During this phase, almost all of the regressions resulted in correlation coefficients (R) exceeding 0.95, indicating the validity of the relationship $F = aN^b$.

While the values of a and b were being studied, it transpired that there was a definite monotonic relationship between the values of a and b. If lognatural (ln) a is now plotted against b, a set of points is obtained between which a straight line is the obvious relationship, as shown in Figure 2. The correlation coefficient (R) for the straight line fit between ln a and b in Figure 2 is 0.989. The 95 percent confidence limits are also shown in Figure 2.

In considering Figure 2, it is obvious that a straight-line relationship between A (= ln a) and B (= b) exists, that is,

$$A = c + mB \quad (1)$$

$$\text{or } \ln a = c + mb \quad (2)$$

It can therefore be stated that a family of linear equations exists:

$$[F_n(x)]_{n\varepsilon\mathcal{N}=(1,2\ldots,65)} = y = A_n + B_n x$$

such that $A_n = c + mB_n$ for every $n \varepsilon \mathcal{N}$

From this it follows that a point $P \varepsilon X \times Y$ exists, which is a solution to the family of linear equations ($y = A_n + B_n x$).

PROOF

For simplicity, consider $\mathcal{N} = (1, 2)$:

The family of linear equations are therefore

$$y = A_1 + B_1 x \text{ and } y = A_2 + B_2 x$$

from which it follows that

$$x = -[(A_1 - A_2)/(B_1 - B_2)]$$

By using

$$A_n = c + mB_n \qquad \text{for } n = 1, 2$$

it is found that

$$x = -\{[c + mB_1 - (c - mB_2)]/(B_1 - B_2)\}$$

$$= -[m\,(B_1 - B_2)/(B_1 - B_2)]$$

$$= -m$$

By substituting this result in $y = A_1 + B_1 x$, it follows that

$$y = A_1 - B_1 m$$

but

$$A_1 = c + mB_1$$

therefore,

$$y = c + mB_1 - mB_1$$
$$= c$$

Therefore

$$P \varepsilon X \times Y \text{ and } P = (-m, c)$$

The coordinates (–m, c) of P are in logarithmic space. If the coordinates in normal space are denoted as (N_o, F_o) it follows that

$$c = \ln F_o$$

and

$-m = \ln N_o$

It then follows that

$$F_o = e^c \quad (3)$$

and

$$N_o = e^{-m} \quad (4)$$

Substituting Equations 3 and 4 in Equation 2 and rearranging:

$$\ln F_o = \ln a + b \ln N_o$$

Therefore

$$F_o = aN_o^b$$

This point at the N_o highest hour, where the hourly traffic volume is a fraction F_o of the AADT, is therefore common to all the counting stations, and is called the focal point.

For the 65 records analyzed, F_o and N_o were determined by applying the relationships in Equations 3 and 4 to the equation of the straight line in Figure 2, resulting in $F_o = 0.072$ and $N_o = 1{,}030$. These values N_o and F_o are the same for all counting stations, and the straight line on the log–log graph of the fraction of AADT versus rank number should pass through this focal point for all counting stations. Using $F = aN^b$ and the relationship $F_o = aN_o^b$, it is easily demonstrated that $F_o = N^b/N_o^b$.

If β is now introduced in the place of b to distinguish between b (calibrated in the initial analysis) and the new value of b (to be calibrated such that $N_o = 1{,}030$ and $F_o = 0.072$), it is found that

$$F/F_o = (N/N_o)^\beta \quad (5)$$

There is therefore only one parameter, β, that determines the peaking characteristics of a given road. (However, the coordinates of the focal point F_o and N_o must be calibrated for a given geographical area.)

If Equation 5 is applied to the data sets, β can be estimated. Figure 3 shows a typical result on normalized scales.

Final Formulation of Peaking Curve

From Equation 5,

$$F/F_o = (N/N_o)^\beta$$

$$F = F_o (N/N_o)^\beta$$

where F is U_n/AADT, and U_n is hourly volume in the nth highest hour (found to be 0.072 and 1,030, respectively, for South African roads).

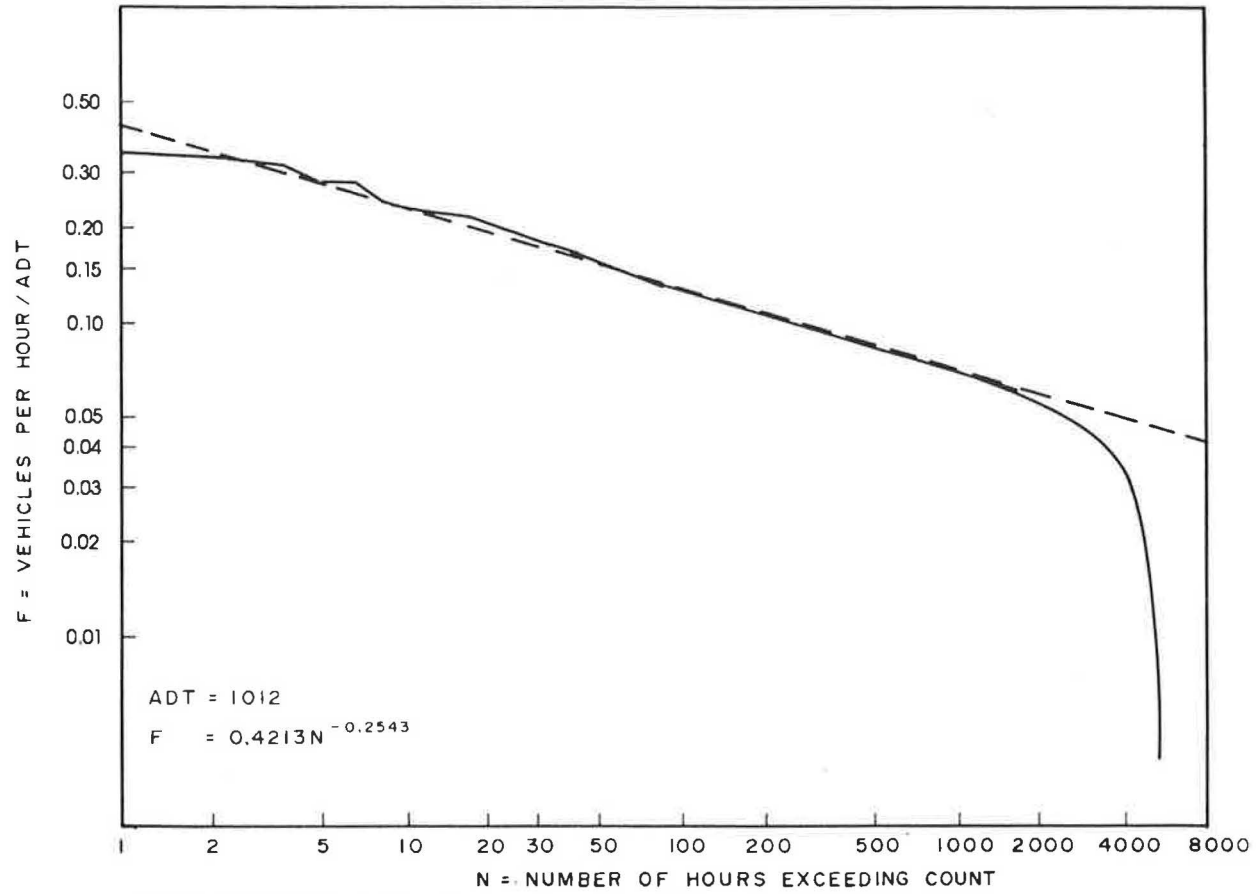


FIGURE 1 Hourly counts and straight-line fit at counting station CH/A, between Christiana and Warrenton, on P3-1.

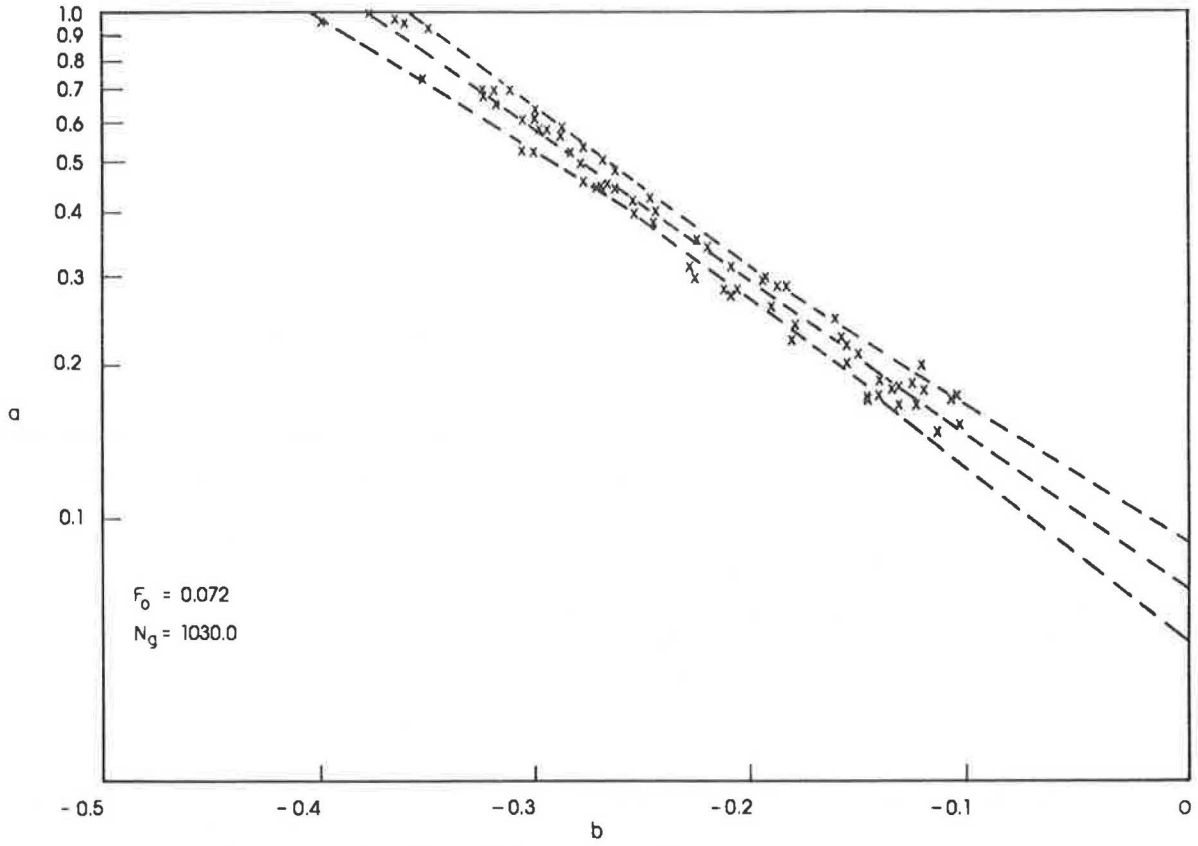


FIGURE 2 Relationship between values of a and b shown on log natural scales.

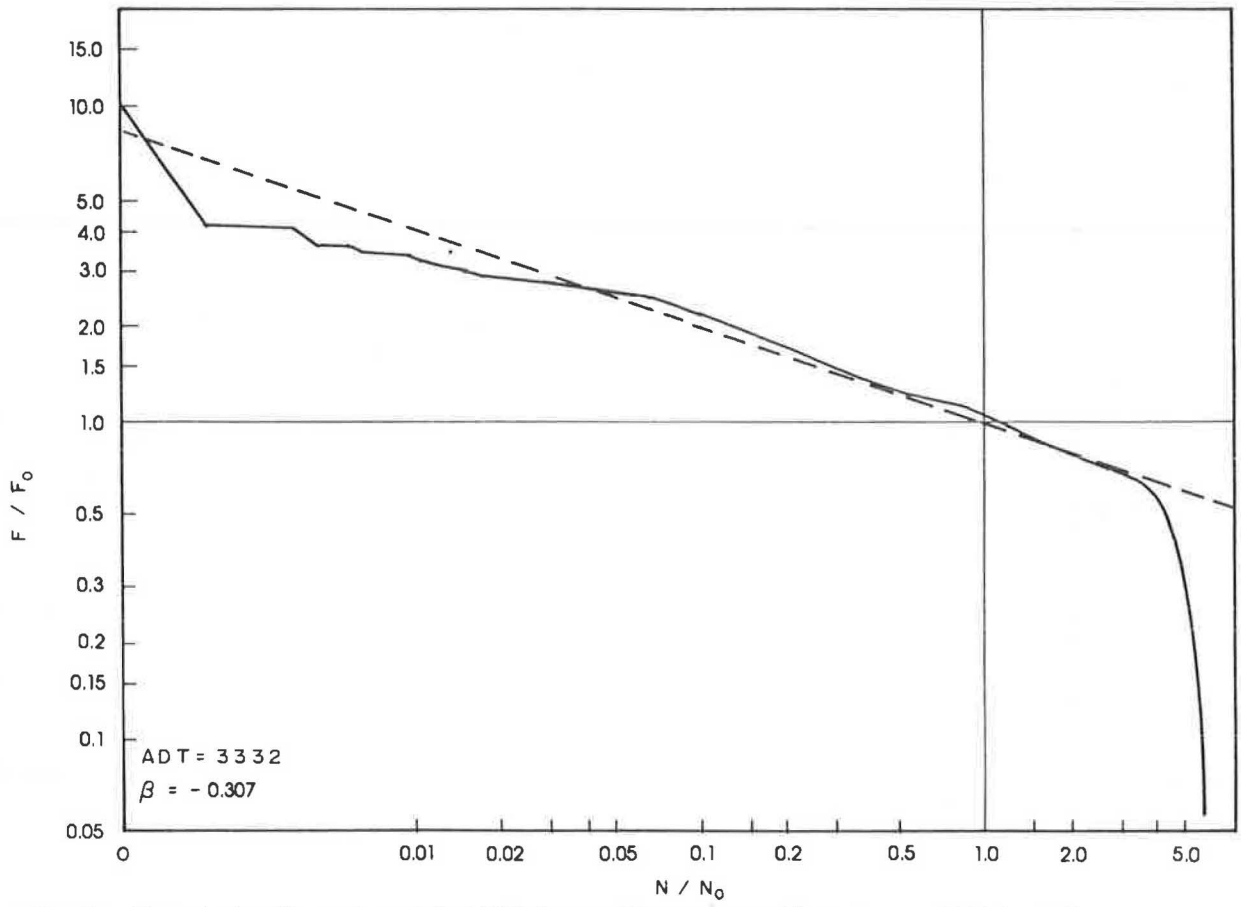


FIGURE 3 Normalized traffic counts at station KG/A, between Krugersdorp and Rustenburg on P16-1, for 1975.

Therefore

$$U_n/AADT = F_o (N/N_o)^{\beta}$$

$$U_n = F_o \cdot AADT \cdot (N/N_o)^{\beta} \quad (6)$$

with

$$F_o = 0.072 \text{ and } N_o = 1,030$$

$$U_n = 0.072 \cdot AADT \cdot (N/1,030)^{\beta} \quad (7)$$

Figures 4 and 5 are graphic illustrations of the model fit. The former shows the fit at a specific location for the 1,000 highest hours and the latter at all stations for the thirtieth highest hour.

SOME PROPERTIES OF PEAKING CURVE

The following relationships can easily be derived:

Number of vehicles in the m highest hours (S_m):

$$S_m = \{(F_o \cdot AADT)/[(\beta + 1) \cdot N_o^{\beta}]\} m^{\beta+1} \quad (8)$$

where $m < N_o$. If the calibrated values $F_o = 0.072$ and $N_o = 1,030$ are used, this equation reduces to

$$S_m = \{(0.072 * AADT)/[(\beta + 1) 1,030^{\beta}]\} m^{\beta+1} \quad (9)$$

where $m < 1,030$.

The fraction of traffic in the m highest hours (F_m):

$$F_m = S_m/(365 * AADT) \quad (10)$$

Number of hours in which the volume exceeds U (N_u):

$$N_u = N_o [U_n/(F_o \cdot AADT)]^{1/\beta} \quad (11)$$

where $N_u < N_o$.

Again, if $F_o = 0.072$ and $N_o = 1,030$ are used,

$$N_u = 1,030 [U_n/(0.072 AADT)]^{1/\beta} \quad (12)$$

where $N_u < 1,030$.

DETERMINATION OF PEAKING CHARACTERISTICS, β

From Continuous Traffic Counts

The hourly counts are ranked, and a normalizing procedure is carried out, that is,

$$y = \text{hourly count}/(AADT \cdot F_o)$$

where $F_o = 0.072$; and

$$X = \text{rank number}/N_o$$

where $N_o = 1,030$.

Now $y = X^{\beta}$ or $\ln y = \beta \ln X$ and a least-squares calibration can be performed to determine β.

BY STUDYING VALUES OF β AT KNOWN LOCATIONS

The frequency and magnitude of the volume of seasonal traffic are the most important determinants of β. If the frequency is high and the magnitude of the traffic volume low, β will tend toward a value of –0.10. On the other hand, if the frequency is low and the magnitude of the traffic volume is high, β will tend toward a value of –0.40.

In selecting a value for β, the following can be used as a guide:

- β = –0.10 indicates a road that has virtually no peaking,
- β = –0.20 is typical of a road that has average peaking, and
- β = –0.40 is typical of a road that has high seasonal peaks.

By Estimating a Value for K, the Fraction of the Thirtieth Highest Hourly Volume

Because engineers may have a historical feel for the K value and may therefore find it easier to estimate this parameter, the following relationship can be used to determine β:

$$\beta = (\ln K - \ln F_o)/\ln (30/N_o) \quad (13)$$

if $F_o = 0.072$ and $N_o = 1,030$, this reduces to

$$\beta = -0.283 \ln K - 0.744 \quad (14)$$

This relationship is demonstrated in the following table.

K (fraction)	β
0.10	–0.092
0.15	–0.207
0.20	–0.289
0.25	–0.352
0.30	–0.403

Because of the mathematical relationship between K and β, the β-methodology would result in exactly the same thirtieth highest hour volume estimate as would the K-methodology. However, the β-methodology allows the estimate of flows in the other ranked hours of the year as well.

From the Relationship Between β and Average Length of Through Trips on a Road

By using the origin–destination matrix together with the assignment routine of the South African Rural Traffic Model, it was possible to determine the average length of through trips on each of the 1,320 links in the model network *(11)*. For those links for which continuous traffic counting data were available,

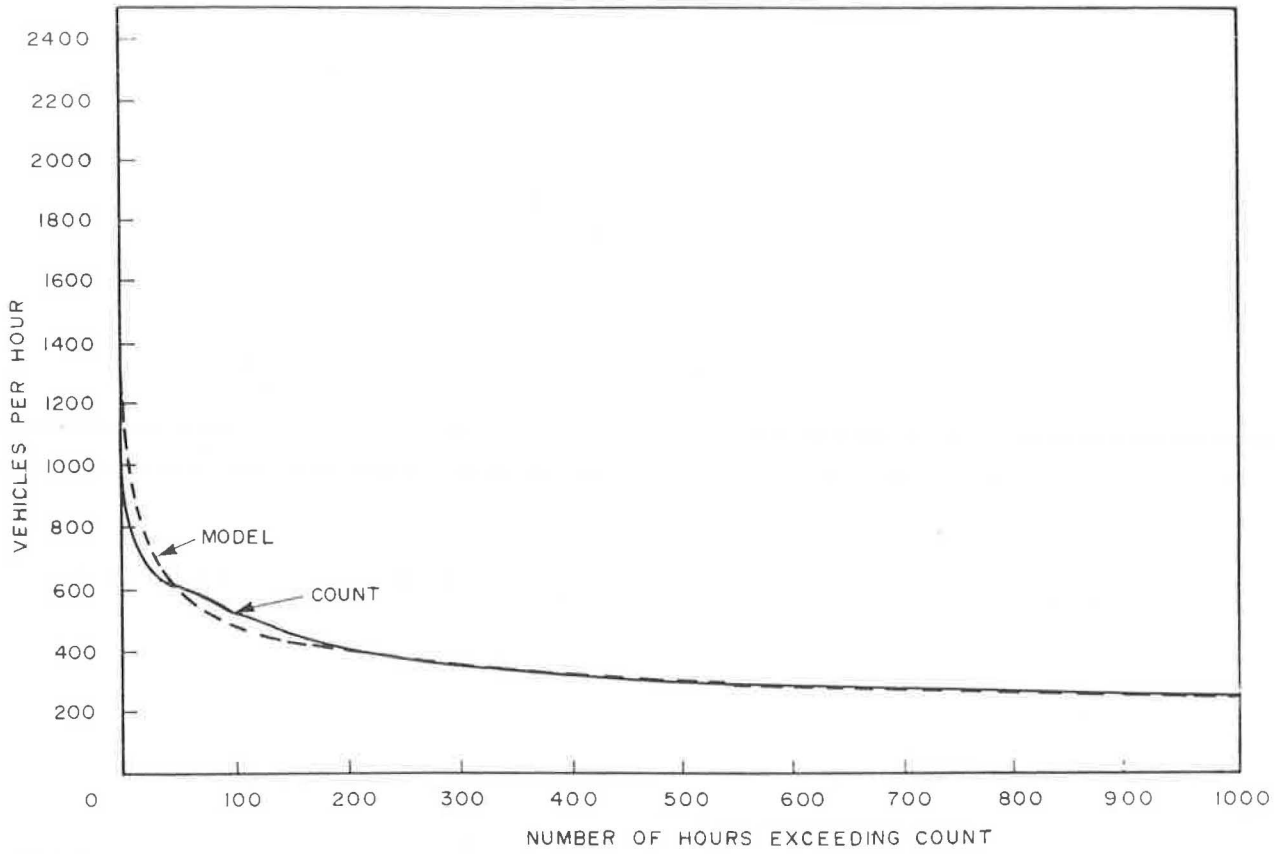


FIGURE 4 Actual counts and modeled traffic at station KG/A, between Krugersdorp and Rustenburg on P16-1.

β was plotted against the average length of through trips. The plot is shown in Figure 6.

The correlation coefficient R for the linear relation shown in Figure 6 is 0.862. The fitted line is given by

$$\beta = 0.0358 - 0.00076L \quad (15)$$

where L is average length of through-trips on the link (km).

Therefore, if the average trip length on a road is known, say from origin–destination surveys or modeling procedures, the peaking characteristic can be estimated. Table 2 gives some trip lengths and associated β-values.

APPLICATIONS OF β-METHODOLOGY

The fact that the hourly flow in any of the 1,000 highest hours can be estimated leads to many new applications. If the flow in the remainder of the hours of the year is assumed to be linearly related to the rank number—which, based on limited samples, is true for the South African situation—the hourly flow in all of the hours of the year can be estimated. In the geometric design process, the following procedures could then be followed.

Phase 1:

- Estimate the base year AADT and β for the road in question.
- Estimate the design year AADT by using growth rates or modeling procedures.
- Estimate the design year β by considering the change in character and/or function of the road.
- Use the β-methodology to determine the hourly flows. This leads to graphs indicating the number of hours in any given year in which the traffic exceeds specified levels of service (LOS). The percentage of traffic that experiences a given LOS and how this changes with time can also be determined. This leads to new insights into the service provided by the proposed facility.

Phase 2:

- Determine the relationship between vehicular operating costs (VOC) and hourly volumes for the specific design alternative.
- Multiply the hourly flow by the VOC for that specific flow, sum over the 8,760 hr of the year, discount to the present for every year of the design life of the facility, resulting in the total discounted VOC on the facility.
- By repeating these two steps for various design alternatives, the differential VOC can be estimated. By comparing this with the differential construction cost, the most economical design can be found.

These procedures, seemingly tedious, are easily performed on modern high-speed computers.

FUTURE RESEARCH

The transferability of the model to the whole of the Republic of South Africa is currently being investigated. Transferability to

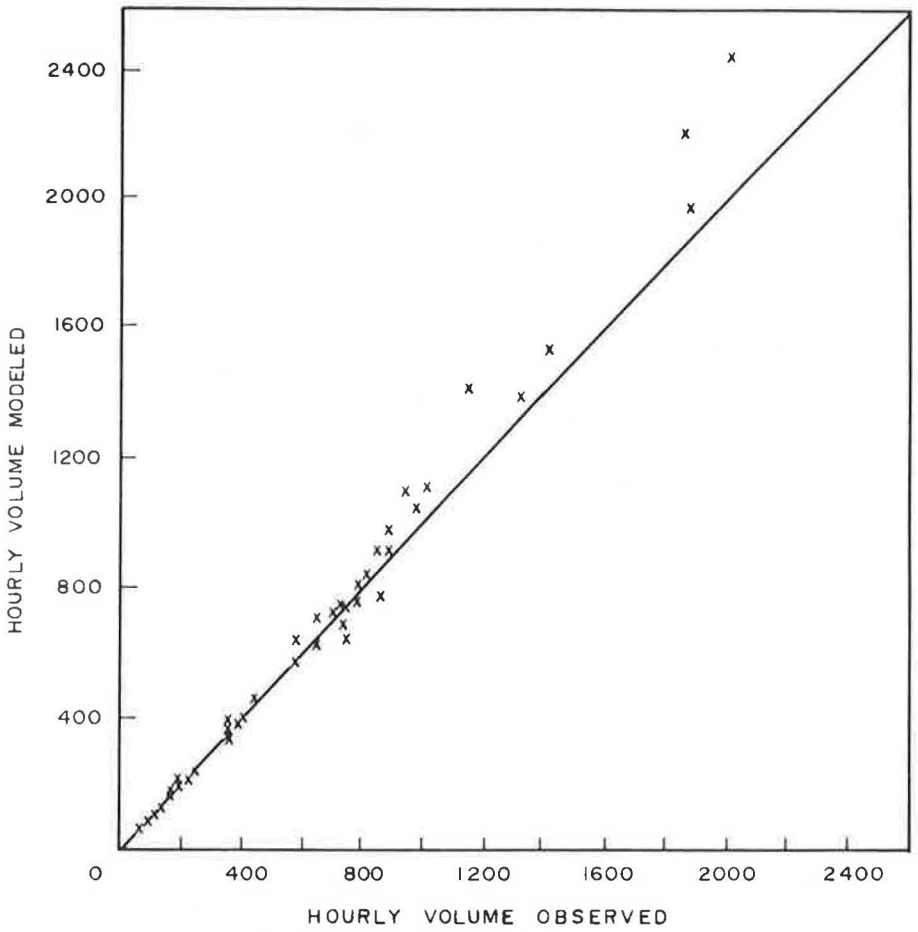


FIGURE 5 Comparison of actual counts and modeled values at the thirtieth highest hour.

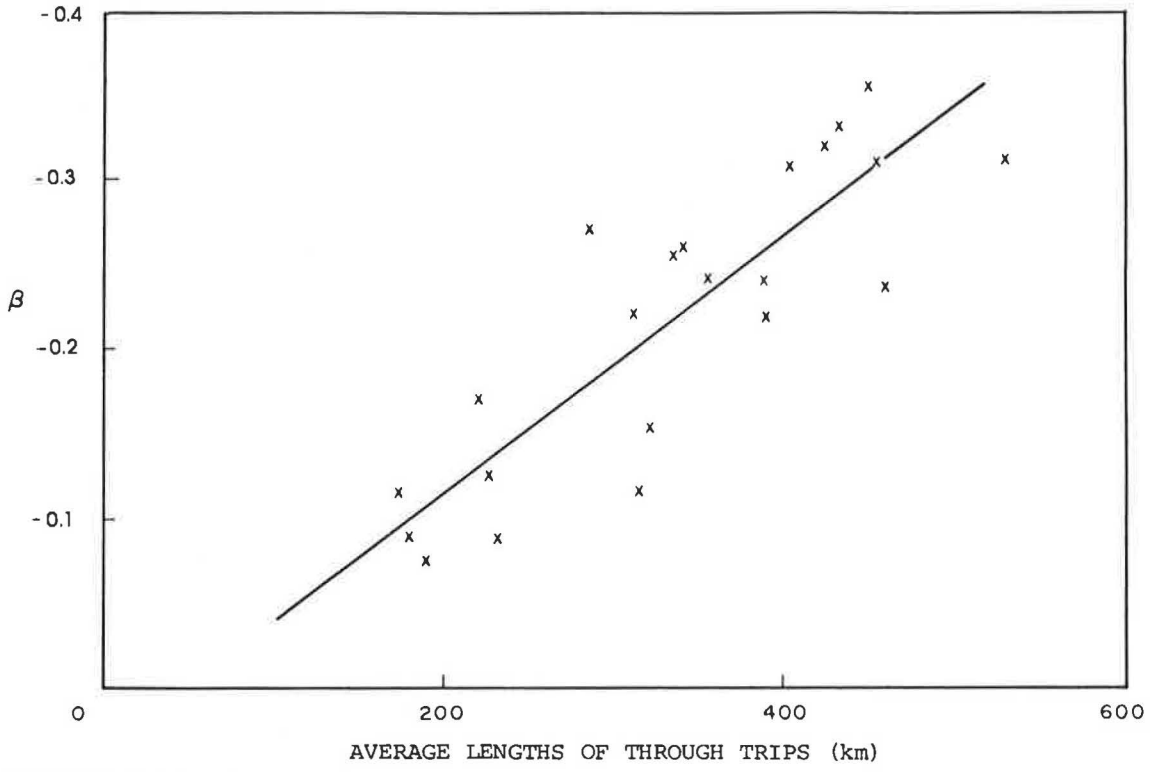


FIGURE 6 Relationship between β and the average length of through trips.

TABLE 2 RELATIONSHIP BETWEEN TRIP LENGTH AND β

Average Trip Length (km)	β	Comments
200	–0.12	Typical of roads with very little peaking
300	–0.19	Typical of the average rural road in South Africa
600	–0.42	Highest peaking roads in South Africa

other countries will in the future also provide interesting insights. The relationship between β and average trip length and the importance of trip length in selecting and classifying traffic counting stations provide avenues for further research. It is also within the bounds of probability that the economic evaluation procedure described in the preceding paragraphs could lead to new insights into the volume concept of design hour.

SUMMARY

A methodology has been developed by estimating the traffic volume in any ranked hour of a year. This methodology depends on no more information than is required for a thirtieth highest hourly volume estimation.

The methodology centers on the normalized relationship that was found:

$$F_n/F_o = (N/N_o)^{\beta}$$

or

$$U_n = F_o \cdot AADT \cdot (N/N_o)^{\beta}$$

The existence of this relationship is of great importance to the road design engineer. It is furthermore believed that this relationship should form the foundation for future research on traffic volume variations and design hour volumes.

ACKNOWLEDGMENTS

This paper is presented with the permission of the Chief Director of the National Institute for Transport and Road Research, Council for Scientific and Industrial Research, South Africa. The contribution of the National Transport Commission toward funding this research is gratefully acknowledged.

REFERENCES

1. L. E. Peabody and O. K. Normann. Applications of Automatic Recorder Data in Highway Planning. *Public Roads*, Vol. 21, No. 11, Jan. 1941.
2. Re-examination of Design Hour Volume Concepts. *ITE Journal*, Vol. 49, No. 9, Sept. 1979, pp. 45–49.
3. *A Policy on Geometric Design of Highways and Streets*. American Association of State Highway and Transportation Officials, Washington, D.C., 1984.
4. W. S. Homburger (ed.). *Transportation and Traffic Engineering Handbook*, 3rd. ed., Prentice Hall, Englewood Cliffs, N.J., 1982.
5. *Geometric Design Standards for Rural Two Lane Two Way Roads*. TMH4. National Institute for Transport and Road Research, Pretoria, South Africa, Sept. 1978.
6. *Growth Rates and 30th Highest Hourly Ratio*. SA Rural Roads Needs Study. Van Niekerk Kleyn & Edwards, Pretoria, South Africa, Nov. 1980.
7. R. F. F. Dawson. *The Economic Assessment of Road Improvement Schemes*. Road Research Technical Paper 75. Her Majesty's Stationery Office, London, England, 1968.
8. G. Phillips and D. Reeson. *Representing the Distribution of Hourly Volumes Through the Year*. TRRL Supplementary Report 804. Transport and Road Research Laboratory, Berkshire, England, 1984.
9. *COBA 9 Manual*. Her Majesty's Stationery Office, Department of Transport, London, England, 1981.
10. *Evaluation of Road User Costs*. Technical Manual for RODES2. Manual P11. CICTRAN. National Institute for Transport and Road Research, Pretoria, South Africa, 1983.
11. P. W. Jordaan, W. W. Crous, and C. A. van Tonder. *The South African Rural Traffic Model: Modelling Concepts and Forecasts*. Vol. H. Annual Transportation Convention, Pretoria, 1984.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Field Validation of Intersection Capacity Factors

JOHN D. ZEGEER

Presented are the results of a series of saturation flow surveys conducted throughout the United States at signalized intersections. The purpose of this research was to verify the saturation flow rates and traffic volume adjustment factors used in various capacity analysis procedures by collecting a relatively extensive data base. Saturation flow headways for more than 20,000 observations were collected for a series of 12 geometric, traffic characteristic, and environmental factors and compared with baseline saturation flow headways for various signal cycle length and phase combinations. Vehicle blockage and lane distribution surveys were conducted for 19,000 additional observations. Based on the results of these surveys, a series of modified adjustment factors is suggested to allow the analyst to determine modified saturation flow rates when calculating signalized intersection capacity.

The signalized intersection chapter of the 1985 *Highway Capacity Manual* (HCM) *(1)* contains a capacity analysis procedure that is based on vehicle delay (stopped delay per vehicle) as the principal measure of effectiveness for levels of service. A critical component used in this procedure is the determination of basic saturation flow rates and adjustment factors used to modify these flow rates.

The purpose of this research effort was to collect an extensive data base that could be used to verify the adjustment factors and the basic saturation flow rate values to improve the reliability of the signalized intersection capacity analysis technique. In addition, these measured saturation flow rates and adjustment factors can be used in other capacity analysis procedures. The saturation flow data collection procedure reported in this paper can be duplicated at other locations.

Results of this research are presented in two sections:

1. Saturation flow rates are provided for the range of common signal cycle lengths and phases, taking into account lost times surveyed.
2. Critical lane volume adjustment factors are presented based on extensive field surveys.

INTERSECTION SATURATION FLOW RATES

Intersection saturation flow rates can be determined for a given level of service based on observed vehicle headways and a design cycle length, as described in Equation 1:

$$SV_E = (3{,}600/h)\ (C - nL)/C \quad (1)$$

where

SV_E = maximum sum of critical lane volumes,
h = average vehicle headway,
C = cycle length,
n = number of phases per cycle, and
L = lost time per phase.

Although this equation takes into account each of the variables that influence the number of vehicles that can be processed by an intersection *(2)*, and this relationship is consistent with the values measured empirically in the Australian signalized intersection capacity method, there are a number of similar formulas that yield reasonably consistent results. [See Table 1, *Transportation Research Circular 212* (*3*, p.7).]

Adjusted critical lane volumes for various cycle lengths and number of phases per cycle can be determined from average vehicle headways and lost times per phase for through lanes.

Typical Saturation Flow Rates

In studies conducted between 1947 and 1979, through-lane saturation flow rates of between 1,500 and 1,800 vehicles per hour of green (vphg) have been reported *(4)*. In these studies, various vehicles in a queue were identified as representing the first vehicle not incurring a significant amount of lost time. The most extensive data base recently reported in the United States was collected in Kentucky in medium and smaller sized communities. An average saturation flow rate of 1,650 vehicles per hour (vph) was found for lanes with widths between 10 and 15 ft and with approach grades between –3 percent and +3 percent *(5)*.

In *Transportation Research Circular 212*, a value of 1,800 passenger cars per hour of green for optimum roadway conditions is recommended as a base value for saturation flows for the fifth vehicle in queue and beyond *(3)*. NCHRP research *(6)* for the 1985 HCM included a review of 2,926 vehicles following the fourth vehicle in queue (contained in film from a 1975 FHWA delay study). A saturation flow rate of 1,827 vphg was measured.

Survey Procedures

A series of surveys of vehicles in saturation flow was conducted throughout the United States at signalized intersection approaches that contained baseline geometric conditions—those approaches not influenced by geometric factors that have been found to significantly reduce saturation flow rates. A description of these conditions is given in Table 1.

Barton–Aschman Associates, Inc., Citicorp Center, 1200 Smith St., Suite 2640, Houston, Tex. 77002.

TABLE 1 BASELINE INTERSECTION APPROACH CONDITIONS

Characteristic	Condition or Value
Lane width	12 ft
Percent heavy vehicles	0 percent (preferred); less than 2 percent (required)
Approach grade	0 percent
Curb parking condition	No curb parking allowed, no illegal curb parking present
Local bus stops	No nearside or farside bus stop activity or signs designating the potential for buses to stop
Area type	Suburban area or outlying commercial; not central business district or residential area
Permitted movements	Straight (or through) movements only; no turn movements permitted in exclusive or optional lane
Metro area size	Greater than 1,000,000 population in metropolitan area
One-way or two-way operation	Two-way operation

Source: Barton–Aschman Associates, Inc.

The data collection techniques used in these surveys were identical to the procedures described in Appendix IV of the 1985 HCM (*1*, Chapter 9). As described in those procedures, the last vehicle stopped in the queue before the onset of green is noted for the signal phase surveyed. At the onset of green, a stopwatch is started. When the rear wheels of the fourth vehicle in queue cross the stop line, saturation flow is considered to begin and the stopwatch time is noted. Saturation flow ends when the rear wheels of the last vehicle (in the queue before the onset of green) cross the stop line. The average saturation flow headway for that signal phase is calculated by dividing the time elapsed for processing n – 4 vehicles by the number of saturation flow vehicles processed. When divided into 3,600, the saturation flow rate for this phase is determined. Only those vehicles crossing the stop line after the fourth queued vehicle were considered to be in saturation flow conditions. In addition, only those vehicles already waiting in a queue at the onset of the green phase were surveyed. Measurements were taken by cycle and by lane at each baseline location.

A total of 3,687 saturation flow vehicles were surveyed at 7 intersection approaches containing baseline conditions in the Chicago, Houston, and Los Angeles metropolitan areas, which was a subset of approximately 20,000 saturation flow headways surveyed. The saturation flow headway for this sample of baseline conditions was 1.92 sec, which is equivalent to a saturation flow rate of 1,875 vphg. Thus, for suburban intersection approaches in large metropolitan areas, saturation flow rates higher than typically reported in the literature can be achieved when good geometric conditions are present. Field surveys at representative intersections (with baseline geometric conditions) should be conducted whenever possible to identify local variations in this value.

Lost Time per Phase

During a green and amber signal phase, there are two periods during which saturation flow headways are not achieved: at the beginning of the phase (startup lost time) and at the end of the phase (clearance lost time).

In an analysis of 1,428 headways for vehicles positioned first, second, or third in queue in Kentucky, the average startup lost time was found to be 1.40 sec (*7*). The largest community surveyed, Louisville, had an average startup lost time of 1.01 sec.

The average startup lost time per phase was determined for the 3,687 vehicles surveyed for the research discussed in this paper on the baseline condition intersection approaches in the Chicago, Houston, and Los Angeles metropolitan areas. The total time elapsed between the onset of green and the time at which the rear wheels of the fourth vehicle in queue crossed the stop line was measured. A value of four times the average saturation flow headway (1.92 sec) was subtracted from the travel time for the first four vehicles to determine the startup lost time for each phase. The average startup lost time thus calculated was 1.31 sec.

The lost time experienced at the end of each phase (clearance lost time) can vary based on the length of the amber phase (clearance phase), width of the intersection, and vehicle approach speed. Most clearance phase lengths range between 3 and 5 sec. However, the clearance lost time rarely is as great as the length of the amber phase. Typically, the clearance loss time (before the onset of green on the cross street) is about one-half of the length of the amber. Clearance lost times at 334 loaded cycles in Kentucky averaged 1.67 sec (*7*). The length of the signal cycle and the change interval (yellow plus all red) had a significant impact on the values observed.

Observations have indicated that the portion of the change interval devoted to an all-red phase is virtually all lost time. Total lost time per phase includes the sum of the average startup lost time and clearance lost time. The total was set as a percent of cycle length. The resulting values of total lost time per phase were used to calculate lane volumes as follows:

- 60-sec cycle—5 percent of cycle—3.0 sec;
- 75-sec cycle—4 percent of cycle—3.0 sec;
- 90-sec cycle—4 percent of cycle—3.5 sec;
- 105-sec cycle—3⅓ percent of cycle—3.5 sec; and
- 120-sec cycle—3 percent of cycle—3.5 sec.

These values of 3.0 and 3.5 sec were selected based on (*a*) the measured startup and clearance lost times reported in the surveys described previously, and (*b*) the recognition that the length of signal cycle does affect the observed value of lost time.

Based on the saturation flow headway and lost time values

just discussed, a series of adjusted critical lane volumes was developed. These values are given in Table 2. It can be observed as the cycle length increases, a slight increase in critical lane volume occurs. Although this may be a valid concept in considering intersection design, the increase in average vehicle delay that is encountered when cycle length is increased is ignored. If the volume on a critical approach is not of sufficient length to fully load a long cycle, the phase will not be used effectively and the flow rate will decrease. Short cycle lengths usually result in lower average vehicle delay for a given set of intersection conditions. Thus, in the intersection design procedure, the shortest cycle length that can accommodate the traffic demands should therefore be considered so that delay is minimized.

As indicated by the data in Table 2, a two-phase signal with a 120-sec cycle length can accommodate 5 percent more critical lane vehicles than would a two-phase signal with 60-sec cycle length. Likewise, increasing the number of phases for a given cycle length results in a reduction in critical lane volume. For example, at a 90-sec cycle length, a four-phase signal can process about 9 percent fewer critical lane vehicles than can a two-phase signal. The critical lane values given in Table 2 should not be extrapolated for cycle lengths of greater than 120 sec because of the increasing likelihood that vehicle platoons will disperse, thus reducing the achievable flow rate.

The adjusted critical lane volumes given in Table 2 were compared with those proposed in *Transportation Research Circular 212* for the signalized intersection operations and design application *(3)*. The values given in *Circular 212* have a threshold value of 1,800 passenger cars per hour for a two-phase signal. Given the 2-sec average vehicle headway assumed for saturation flow conditions in that procedure, the saturation flow values in *Circular 212* do not take into account the effect of lost time on saturation flow. The three- and four-phase values in *Circular 212* reflected 4 percent and 8 percent reductions in critical volume, respectively. These reductions are consistent with the reductions in critical lane volumes given in Table 2 in this paper for a 90-sec cycle length.

ADJUSTMENT FACTORS

The adjustment factors used to modify saturation flow rates vary significantly by source. Many of the factors from other sources have been estimated, without the benefit of research. Other factors have been based on a very small sample of data. Still other factors appear to rely on the 1965 HCM values *(8)*, which were derived from surveys conducted in the late 1950s. A major portion of the research effort documented in this paper involved the review of adjustment factors proposed in recent literature. The results from this new data collection effort were then used to verify the reliability of the adjustment factors contained in *Transportation Research Circular 212 (3)* and the 1985 HCM *(1)*.

The 12 factors that are most commonly considered to affect saturation flow rates were analyzed as a part of the research effort. These factors are as follows:

1. Lane width,
2. Heavy vehicles,
3. Vertical grade on approach,
4. Curb parking,
5. Local bus stop activity,
6. Area type,
7. Through-lane utilization,
8. Turning movements (single turn lanes),
9. Dual turn lanes,
10. Pedestrian conflict,
11. Metropolitan area size, and
12. One-way or two-way operation.

All of the surveys except for numbers 5, 7, 9, and 10 used the saturation flow technique as described in Appendix IV of the 1985 HCM (*1*, Chapter 9), and described in a previous section of this paper. The survey of Factor 5 and a portion of the survey of Factor 4 (Parking and Unparking Maneuvers) were conducted by observing moving lane blockage time experienced during green signal phases. The surveys of Factors 7 and 9 were conducted by recording approach vehicle volumes distributed by lane. Values for Factor 10 (and permissive left-turn values for Factor 8) were calculated based on surveys reported in other sources.

Surveys were conducted in 14 metropolitan areas throughout the United States at a total of 98 signalized intersection approaches. More than 20,000 vehicles operating in saturation flow conditions were surveyed to determine average vehicle headways for the various factors (Factors 1, 2, 3, 4, 6, 8, 11, and 12). To analyze each individual factor, intersections were

TABLE 2 CRITICAL LANE VOLUMES

Cycle Length (sec)	Maximum Sum of Critical Lane Volumes[a] (passenger cars/hr)		
	Two-Phase Signal	Three-Phase Signal	Four-Phase Signal
60	1,690	1,590	1,500
75	1,720	1,650	1,570
90	1,730	1,660	1,580
105	1,750	1,690	1,630
120	1,770	1,710	1,660

Source: Barton–Aschman Associates, Inc.

[a]Sum of critical lane volumes: $CV = (3{,}600/C)\ (C - nL/h) = (3{,}600/h)\ (C - nL/C) = (3{,}600/h)\ (1 - nL/C)$.

where C = cycle length; n = number of phases per cycle; L = lost time per phase—varies between 3.0 and 3.5 sec; h = saturation flow headway = 1.92 sec; and $CV = (1{,}875)\ (1 - nL/C)$.

selected that contained all of the baseline geometric conditions, except for the factor being tested. The results from the saturation flow surveys were then compared with the baseline saturation flow values to isolate the effect of the factor in question. More than 19,000 additional vehicles were counted to determine lane distribution rates, parking and unparking maneuvers, and bus stop maneuver influence (Factors 4, 5, 7, and 9).

Each of the 12 factors is discussed separately in the following subsections. Adjustment factors contained in *Transportation Research Circular 212 (3)* and the 1985 HCM *(1)* are reviewed and compared with the results of the data surveys.

Lane Width

The 1985 HCM adjustment factors for lane width are graduated in foot-by-foot increments between 8 and 15 ft *(1)*, as given in Table 3.

TABLE 3 ADJUSTMENT FACTORS FOR LANE WIDTH AS GIVEN IN THE 1985 HCM

Lane Width (ft)	f_{LW}
8	0.87
9	0.90
10	0.93
11	0.97
12	1.00
13	1.03
14	1.07
15	1.10
16	Use 2 lanes

The *Transportation Research Circular 212* review of lane width adjustment factors from various sources concludes that, "Lane widths in the 10- to 13-foot range have little effect on saturation flow or capacity" *(3)*. Thus, a step function is recommended whereby passenger car volumes are increased 10 percent for lane widths of between 8.0 and 9.9 ft and decreased by 10 percent for lane widths of between 13.0 and 15.9 ft (compared with a fixed saturation flow rate). In the Australian capacity procedures, capacity adjustments are made for lane widths outside the range of 10.0 to 12.0 ft, also using step functions.

As a part of the national surveys conducted for the research documented in this paper, saturation flow rates were measured on 11 approaches with lane widths varying between 8.5 and 9.5 ft. The sample size was 2,733 saturation flow vehicles. Four approaches with lane widths varying between 13.0 and 15.5 ft were surveyed, with a sample size of 1,568 saturation flow vehicles. All baseline conditions except for lane width were held constant at these locations. The survey results were then compared with those of the baseline condition surveys (with a sample size of 3,687 saturation flow vehicles). The narrower lane widths demonstrated saturation flow rates between 2 and 5 percent less than did those in the baseline surveys, while the wider lane widths demonstrated saturation flow rates 5 percent greater than did those in the baseline surveys. Thus, the following factors are proposed for lane-width adjustments based on the survey findings:

- 8 to 8.9 ft—0.95;
- 9 to 9.9 ft—0.98;
- 10 to 12.9 ft—1.00; and
- 13 to 15.9 ft—1.05.

Heavy Vehicles

A heavy vehicle is defined as any truck or bus having six or more tires on the pavement. All vans and light-duty trucks containing only four tires are excluded from this definition of heavy vehicles. Data used as input for the 1985 HCM indicated that the headway between preceding passenger vehicles and trucks averaged 2.06 sec (108 vehicle samples) *(6)*. The headway between trucks and following passenger vehicles averaged 2.61 sec (105 vehicle samples). Average vehicle headways between all vehicles preceding or following a bus were 3.10 sec (30 vehicle samples). This implies a heavy vehicle passenger car equivalent of between 1.3 and 1.6. In *Transportation Research Circular 212*, it is recommended that a passenger car equivalent value of 2.0 for all heavy vehicles (trucks or through buses not stopping at the intersection) be used to convert from vehicles to passenger cars. Surveys conducted by Carstens in 1971 for trucks in through lanes resulted in a heavy vehicle-passenger car equivalent of 1.6 *(9)*.

The surveys conducted for the research documented in this paper at intersections with significant volumes of heavy vehicles on the approaches indicated that heavy trucks had a significant influence on increasing vehicle headways. First, headways of automobiles following automobiles were surveyed to confirm that their saturation flow rates were comparable to baseline conditions (1.92-sec headways). Then, surveys were conducted to determine average headway values for three conditions: trucks following automobiles (68 samples)—5.16 sec; automobiles following trucks (64 samples)—2.22 sec; trucks following trucks (34 samples)—3.76 sec. These data suggest a passenger car equivalent of 1.92, which results in the proposed adjustment factors for heavy vehicles given in Table 4.

Vertical Grade on Approach

The headway data collected by intersection approach grade (percent) during the NCHRP study leading to the preparation of the 1985 HCM generally showed reductions in saturation flow rate for both upgrades and downgrades *(6)*. The reductions found in the NCHRP study for upgrades were as follows: 31 percent for 3 percent upgrades (55 vehicles sampled), 4.6 percent for upgrades of between 4 and 5 percent (38 vehicles sampled), and 23.9 percent for upgrades of between 6 and 7 percent (70 vehicles sampled). The surveys excluded heavy vehicles. The larger difference in saturation flow rate for 3 percent upgrades compared with that of steeper upgrades was noted. A reduction was also generally found to occur in saturation flow rates for downgrades: a 9.6 percent decrease in flow rates for downgrades of 3 percent (234 vehicles sampled), a slight (0.5 percent) increase in flow rates for downgrades of

TABLE 4 PROPOSED ADJUSTMENT FACTORS FOR HEAVY VEHICLES

Percent Heavy Vehicles	f_{HV}
0	1.00
2	0.98
4	0.96
6	0.95
8	0.93
10	0.92
20	0.84
30	0.78

between 4 and 5 percent (243 vehicles sampled), and a decrease of 10.2 percent in flow rates for downgrades of between 6 and 7 percent (24 vehicles sampled). An increase in flow rates for downgrades and a decrease in flow rates for upgrades are recommended in the 1985 HCM *(1)*. The suggested 1985 HCM factors are given in Table 5.

TABLE 5 SUGGESTED 1985 HCM FACTORS FOR VERTICAL APPROACH ON GRADE

Percent Grade	f_G
–6	1.03
–4	1.02
–2	1.01
0	1.00
+2	0.99
+4	0.98
+6	0.97

A sample of 1,592 saturation flow vehicles was surveyed on approaches with upgrades of between 4 and 6 percent as a part of the field surveys for the research documented in this paper. On downgrades of between 4 and 6 percent, a total of 1,697 saturation flow vehicles were surveyed. These surveys showed reductions in flow rates of 4 percent for downgrades and 10 percent for upgrades on intersection approaches. Observations at these intersections suggested that drivers on relatively steep downgrades proceeded through the intersections in a manner that was somewhat more tentative than that used on intersection approaches with no discernable grade. This resulted in an increase in headway between vehicles and a resulting decrease in saturation flow rates. On the basis of the results of these surveys, the following factors are proposed for intersection approach grades:

- 5 to 6 percent downgrade—0.96;
- 3 to 4 percent downgrade—0.98;
- 2 percent downgrade to 2 percent upgrade—1.00;
- 3 to 4 percent upgrade—0.95; and
- 5 to 6 percent upgrade—0.90.

Curb Parking

Research for the 1985 HCM used FHWA films taken in 1975 of vehicle parking movements *(6)*. It was indicated in this research that saturation flow rates were 9.2 percent lower in lanes adjacent to curb parking than in similar lanes not adjacent to curb parking (229 headways surveyed). In *Transportation Research Circular 212*, it is stated that, "Most North American techniques do not explicitly consider a reduction in capacity due to parking, if the parking ends 250 feet before the intersection stops. . ." *(3)*.

A sample of 1,811 saturation flow vehicles was collected as a part of the field survey for the research documented in this paper. The results of this survey indicated that saturation flow rates were reduced by 11 percent for vehicles traveling in lanes adjacent to curb parking (where parked vehicles exist between the stopline and 250 ft behind the stopline). A second series of surveys was conducted for parking and unparking maneuvers to determine the amount of time that the adjacent moving lane was blocked during these maneuvers. The average maneuver blocked the adjacent lane for about 7 sec (178 samples). Assuming that all parking or unparking maneuvers occur during the green signal phase on the approach being analyzed (because standing vehicles during the red phase would block the ability of a vehicle to enter or leave a parking space), the saturation flow time available for vehicles on an approach with one moving lane would be reduced by 3.9 percent for each 10 parking or unparking maneuvers per hour within 250 ft of the stopline. On this basis, the factors given in Table 6 are proposed for approaches with curb parking activity. These factors are similar to, but not exactly the same as, those contained in the 1985 HCM.

Local Bus Stop Activity

Research for the 1985 HCM included the review of 12 buses that stopped at traffic signals, blocking traffic from entering the intersection during the green phase *(6)*. The average length of

TABLE 6 FACTORS PROPOSED FOR CURB PARKING ACTIVITY

No. of Lanes on Approach	No Parking	No. of Parking Maneuvers/Hr				
		0	10	20	30	40
1	1.00	0.89	0.86	0.82	0.79	0.75
2	1.00	0.94	0.93	0.91	0.89	0.87
3	1.00	0.97	0.96	0.95	0.94	0.93

time for each bus to load and unload passengers was 20.7 sec. The average time that vehicles were blocked during a green phase was 14.0 sec. There were 2.4 vehicles blocked on average while a bus was entering or leaving the bus stop. Observations of the 12 buses were used to calibrate bus blockage adjustment factors for the 1985 HCM. An average passenger car equivalent value of 5.0 was suggested in *Transportation Research Circular 212* for local buses *(3)*. This is equal to an average headway of about 10 sec per bus (including those that stop at the intersection and those that do not stop).

The surveys for the research documented in this paper included observations of 262 buses stopping on intersection approaches. The average length of time that these buses blocked the adjacent lane during a green phase was 9.1 sec (compared with 14.0 sec from the 1985 HCM surveys). This resulted in a reduction of saturation flow of about 2.5 percent for each of 10 buses per hour stopping on a one-lane approach. The resulting bus blockage factors based on this data base, to be proposed, are given in Table 7.

TABLE 7 PROPOSED RESULTING BUS BLOCKAGE FACTORS

No. of Lanes on Approach	No. of Buses Stopping/Hr				
	0	10	20	30	40
1	1.00	0.97	0.95	0.92	0.90
2	1.00	0.99	0.97	0.96	0.95
3	1.00	0.99	0.99	0.98	0.97

Area Type

The influence of the surrounding environment on service flow rates is commonly taken into account by defining the intersection location within a metropolitan area as residential (RES), outlying commercial district (OCD), or central business district (CBD). Research for the 1985 HCM identified vehicle headway values of 2.11 and 2.09 sec for OCD and RES locations, respectively *(6)*. The sample sizes for these two area type surveys were 781 and 187 vehicles, respectively. Headway values in CBD locations were found to average 2.35 sec (500 vehicles sampled). As a result, an adjustment factor of 0.9 for intersections within CBDs is suggested in the 1985 HCM *(1)*.

The surveys for the research documented in this paper included sample sizes of 2,687, 709, and 883 saturation flow vehicles for OCD, RES, and CBD locations, respectively. An analysis of the survey results indicated that the saturation flow rates for vehicles in RES areas are 1 percent greater than the saturation flow rates in OCD areas, and that the saturation flow rates in CBD areas are 1 percent less than the saturation flow rates in OCD areas. Thus, no adjustment factors are proposed for area type.

Through-Lane Utilization

Lane utilization factors take into account that vehicles in exclusive through lanes do not distribute equally among the lanes available on an approach. A review of FHWA films by the NCHRP study team (conducting research for the 1985 HCM) identified a mean value for the high volume in two through lanes as 54 percent. The mean value for the high volume in three through lanes was 39 percent. These percentages result in lane utilization factors of 1.08 and 1.17 for two and three through lanes, respectively. In the 1985 HCM, lane utilization factors of 1.05 for two-lane approaches and 1.10 for three or more lane approaches are recommended (reductions of 5 percent and 9 percent of capacity). In *Transportation Research Circular 212*, a lane utilization factor of 1.05 for a two-lane approach (which represents 52.5 percent of the approach volume in the heavier lane) and a 1.10 factor for a three-lane approach (which represents 37 percent of the approach volume in the heaviest lane) are recommended.

Surveys of 5,600 vehicles on two-lane approaches and 6,900 vehicles on three-lane approaches were conducted throughout the county for the research documented in this paper. The results of these surveys indicated that 51.9 percent of the through volume occurred in the heavier lane on a two-lane approach and 36.5 percent of the through volume occurred in the heaviest lane on a three-lane approach. Based on the survey results, lane utilization factors of 0.96 and 0.91 are proposed for approaches with two lanes and three or more lanes, respectively. These values are consistent with those specified in the 1985 HCM *(1)*.

Turning Movements

Exclusive Turn Lanes—Protected Phase

Left-Turn Lanes A reduction of 15 percent in saturation flow rates for exclusive left-turn lanes with protected signal phases was found in the NCHRP study *(6)*. This was based on a sample size of 205 left-turning vehicles (in saturation flow after the fourth vehicle in queue). The adjustment factor contained in the 1985 HCM for this condition is 0.95, which compares with a 3 percent reduction recommended by Messer and Fambro in their paper titled "Critical Lane Analysis for Intersection Design" *(2)*. In *Transportation Research Circular 212*, the recommended adjustment factor for an exclusive left-turn lane with a protected turn phase is 1.05 (equivalent to a 5 percent reduction in capacity) *(3)*.

Results of surveys of 774 left-turning vehicles in saturation flow conducted for the research documented in this paper indicated an average 3 percent reduction in saturation flow rate when compared with through-lane headways at comparable locations. Thus, a 0.97 factor for single left-turn lanes with protected phases is proposed.

Right-Turn Lanes The 1985 HCM includes a right-turn adjustment factor of 0.85 for exclusive right-turn lanes controlled by protected signal phases *(1)*. Messer and Fambro recommend that "when a separate right-turn lane is provided, neither right-turning volume nor right-turn lane is analyzed" *(2)*. Surveys were conducted for the research documented in this paper for 723 right-turning vehicles in saturation flow. The locations surveyed had curb radii of between 10 and 30 ft. It

was found that the average saturation flow headway for these right-turning vehicles was 19 percent longer than through-vehicle headways at comparable locations. Thus, a factor of 0.84 is proposed for exclusive right-turn lanes with short turn radii. When turn radii of between 25 and 44 ft were provided, the Kentucky research found a saturation flow rate that was 8 percent higher than that for narrow radii *(5)*. Thus, a factor of 0.91 is proposed for wider right-turn radii (30 to 44 ft).

Exclusive Turn Lanes—Permissive Phase

Left-Turn Lanes The 1985 HCM includes a procedure for exclusive left-turn lanes controlled by permissive signal phases that involves 11 calculations to determine an adjustment factor. In *Transportation Research Circular 212*, the recommended adjustment factor for an exclusive left-turn lane that operates on a permissive phase is based on the through and right-turn opposing volume *(3)*. The factors (expressed in passenger car equivalents) range between 1.0 for opposing volumes of 0 to 299 and 6.0 for opposing volumes greater than 1,000.

The field studies conducted in Kentucky for lost time included a determination of the relationship between lost time per cycle and opposing volume per cycle for opposed left turns *(7)*. This relationship was based on surveys conducted at locations where left-turning vehicles move across one lane of opposing traffic (a two-lane street) and across two lanes of opposing traffic (a four-lane street). The survey results from Kentucky were analyzed for this paper and a set of adjustment factors was derived by calculating the incremental lost time per cycle (at various opposing volumes) that would result from the additional delay experienced by the left-turning vehicles. This incremental value of lost time was then converted to adjustment factors for a range of opposing volumes. The proposed adjustment factors are given in Table 8.

Right-Turn Lanes The two primary factors that influence the saturation flow rates for right-turning vehicles operating in an exclusive lane without a protected phase are the right-turn cornering radius and the extent of pedestrian interference in the opposing crosswalk. Thus, the volume in this lane should be factored by the appropriate radius factor (either 0.84 or 0.91, as previously described) and a pedestrian blockage factor, as presented in a subsequent section. This proposed procedure is consistent with that included in the 1985 HCM *(1)*.

TABLE 8 ADJUSTMENT FACTORS FOR EXCLUSIVE LEFT-TURN WITH PERMISSIVE PHASE

Opposing Volume	Adjustment Factor	
	Two-Lane Street	Four-Lane Street
100	0.94	0.98
200	0.87	0.96
300	0.81	0.94
400	0.75	0.92
500	0.69	0.89
600	0.63	0.87
700	0.58	0.85
800	0.53	0.83
900	0.47	0.81
1,000	0.41	0.80
1,100	0.35	0.79
1,200	0.30	0.76

Optional Turn–Through Lanes—Protected Phase

The 1985 HCM does not distinguish between the adjustment factor for a left-turning or right-turning vehicle in an exclusive lane or in an optional turn–through lane as long as the turn movement operates during a protected phase. In both cases, the 1985 HCM adjustment factor is dependent on the base factor appropriate for an exclusive turn lane adjusted for the proportion of turning vehicles in the shared lane *(1)*. A passenger car equivalent value of 1.2 for optional turn–through lanes with protected signal phases is suggested in *Transportation Research Circular 212 (3)*. This is equivalent to a 17 percent reduction in saturation flow rate.

Consistent with the 1985 HCM procedure, it is proposed that either the left-turn saturation flow factor (0.97) or the appropriate right-turn saturation flow factor (0.84 or 0.91) be modified by the percent of turning vehicles in the lane. For example, the left-turn adjustment factor should vary as follows:

Percent of Left Turns in Shared Lane	*Adjustment Factor*
100	0.97
67	0.98
33	0.99
0	1.00

Optional Turn–Through Lanes—Permissive Phase

Left-Turning Vehicles For left-turning vehicles in an optional turn–through lane that do not have a protected signal phase, the 1985 HCM uses the same procedure as for the permissive phase—exclusive left-turn lane option—with a modification introduced to recognize the portion of left-turning vehicles in the shared lane *(1)*. The factors in *Transportation Research Circular 212* for left-turning vehicles in a shared lane are identical for both the protected and permissive phase options *(3)*.

The lost time surveys conducted in Kentucky allowed for a relationship to be determined between the change in lost time per cycle for a range of opposing volumes and the percent of left-turning vehicles in the shared lane *(7)*. These survey results were analyzed for this paper to derive a set of adjustment factors for this shared lane–permissive phase option, consistent with the factors developed for the exclusive lane–permissive phase option. Table 9 gives the proposed adjustment factors for a range of opposing volume values. Note that in this table the opposing flow is expressed in volume per lane rather than total opposing volume (as in Table 8).

Right-Turning Vehicles Factors for right-turning vehicles operating in a shared lane–permissive phase option should be

TABLE 9 LEFT-TURN ADJUSTMENT FACTORS FOR SHARED THROUGH–LEFT-TURN LANE WITH PERMISSIVE PHASE

Opposing Volume/Lane	Percent Left-Turn Volume in Shared Lane		
	10	30	50
100	0.99	0.98	0.97
200	0.99	0.96	0.95
300	0.98	0.94	0.92
400	0.97	0.92	0.91
500	0.96	0.90	0.88
600	0.95	0.88	0.86
700	0.93	0.86	0.82
800	0.91	0.84	0.80
900	0.90	0.80	0.77
1,000	0.87	0.78	0.75
1,100	0.85	0.76	0.71
1,200	0.83	0.72	0.70

based on the factors derived for right-turning vehicles operating in a shared lane–protected phase option. The appropriate right-turn saturation flow factor (0.84 or 0.91) should be modified by the percent of turning vehicles in the shared lane. For example, it is proposed that the right-turn adjustment factors for an optional lane with a 10- to 30-ft right-turn radius be as follows:

Percent of Right Turns in Shared Lane	*Adjustment Factor*
100	0.84
75	0.88
50	0.92
25	0.96
0	1.00

Both the right-turn and left-turn volumes in shared turn–through lanes should also be adjusted by the pedestrian conflict adjustment factor described in a subsequent section.

Dual Turn Lanes

A reduction factor for dual exclusive turn lanes reflects the potential for uneven distribution of vehicles in the two turn lanes. The 1985 HCM includes adjustment factors of 0.92 for dual left-turn lanes and 0.75 for dual right-turn lanes *(1)*. It is suggested in *Transportation Research Circular 212* that 55 percent of the total left-turn volume be assigned to one lane for a dual left-turn lane. This results in an adjustment factor of 0.91 for the dual left-turn lane volume. Stokes found in his surveys in Texas that 50.6 percent of vehicles in dual left-turn lanes (sample size of 3,458) used the outside lane *(4)*. This results in an adjustment factor of 0.97.

Surveys were also conducted for the research documented in this paper for dual exclusive turn lanes to determine volume distribution. The sample sizes for these surveys were 2,026 and 856 for dual left-turn and dual right-turn lanes, respectively. A total of 50.3 percent of the vehicles surveyed in dual left-turn lanes were in the curb (inside) lane and 55.3 percent of the vehicles surveyed in dual right-turn lanes were in the curb (inside) lane. This was a significantly high use of the curb lane for dual right-turn movements because the right-turning radii varied from 10 to 50 ft. The lane distributions suggest factors of 0.97 and 0.90 for the dual left-turn and dual right-turn lanes, respectively, when compared with single exclusive turn lanes. Combining the single and dual turn lane adjustment factors for protected turn phase operation yields an overall factor that can be applied to the unadjusted turn volumes in dual exclusive turn lanes (yielding one adjustment factor rather than two for each dual turn lane calculation). The recommended overall adjustment factors are as follows: dual left-turn lane—0.94, and dual right-turn lane—0.76.

Pedestrian Conflict

Pedestrians in the opposing street crosswalk conflict with both right-turning and left-turning vehicles when they must turn without a protected phase. It is suggested in the 1985 HCM that right-turn volumes be adjusted by a factor that takes into account the number of pedestrians per hour in the conflicting crosswalk but does not suggest an adjustment for permissive left turns. This adjustment reduces the basic right-turn factor (0.85) in proportion to the conflicting pedestrian volume up to a maximum of 1,700 pedestrians per hour. In *Transportation Research Circular 212*, a series of passenger car equivalent values for right-turning vehicles is proposed based on pedestrian volume ranges in the conflicting crosswalk. These values are 1.0 for pedestrian volumes under 100 per hour, 1.25 for pedestrian volumes between 100 and 600, 1.50 for pedestrian volumes between 600 and 1,200, and 2.0 for pedestrian volumes greater than 1,200.

A detailed survey was conducted for UMTA by others in Washington, D.C. *(9)*. As a result of these surveys, the relationship between additional vehicle delay (beyond that delay normally experienced when turning) and the number of pedestrians per cycle in the crosswalk was developed. These survey results were used for the research documented in this paper to derive an equivalent percent reduction in green time available for right-turning vehicles in relation to the number of pedestrians per hour for various signal cycle lengths. The resulting adjustment factors for pedestrian flows (expressed in terms of the number of pedestrians per hour) are given in Table 10.

TABLE 10 RESULTING ADJUSTMENT FACTORS FOR PEDESTRIAN FLOWS

No. of Pedestrians/Hr	Factor	No. of Pedestrians/Hr	Factor
100	0.93	1,100	0.75
200	0.91	1,200	0.73
300	0.89	1,300	0.71
400	0.87	1,400	0.68
500	0.86	1,500	0.66
600	0.84	1,600	0.65
700	0.82	1,700	0.63
800	0.80	1,800	0.61
900	0.78	1,900	0.59
1,000	0.76	2,000	0.57

The values in Table 10 provide a less significant adjustment factor at the low end of pedestrian volumes (100 to 600 pedestrians) than do the *Transportation Research Circular 212* factors *(3)*. A reduction of between 7 and 16 percent is proposed for low pedestrian volumes versus a 20 percent reduction for comparable volumes contained in the *Circular 212* technique. It is suggested that the adjustment factors given in Table 10 be applied to the portion of right- and left-turning vehicles turning from exclusive or optional turn lanes on permissive signal phases.

Metropolitan Area Size

The *Transportation Research Circular 212* analysis technique *(3)* and the 1985 HCM *(1)* do not take into account variations in driver characteristics based on city size. Data from two medium-sized cities (population of from 250,000 to 400,000) and from two large cities (population of more than 1,000,000) were reviewed as a part of the NCHRP study *(6)*. The results of this data review indicated that saturation flow rates in the two medium-sized cities were 11 percent higher than saturation flow rates in the two large cities.

Surveys in Kentucky were conducted in two communities with populations of more than 100,000 persons, three communities with populations of between 20,000 and 50,000 persons, and three communities with populations of less than 20,000 persons *(5)*. Saturation flow rates in the cities with 20,000 to 50,000 persons were 8 percent lower than in the largest city surveyed. Saturation flow rates in the cities with populations less than 20,000 persons were 17 percent lower than those in the largest city surveyed.

The surveys for the research documented in this paper for the metropolitan area size factor included an analysis of saturation flow rates in three communities with populations of between 50,000 and 100,000 persons and three communities with populations of between 300,000 and 800,000 persons. The sample sizes for these surveys were 671 saturation flow vehicles and 1,169 saturation flow vehicles, respectively. The surveys were compared with the baseline saturation flow surveys, collected at intersections in metropolitan areas with populations greater than 1,000,000 persons. The results of the data analysis indicated that saturation flow rates in communities with populations of between 300,000 and 800,000 persons were identical to saturation flow rates in larger communities. Saturation flow rates in communities with populations of between 50,000 and 100,000 persons were 9 percent lower than in the other communities surveyed (consistent with the Kentucky data for the population category of 20,000 to 50,000). Thus, the following adjustment factors are proposed to reflect the effect of metropolitan area size: population of more than 100,000—1.00; population of between 20,000 and 100,000—0.91; and population of less than 20,000—0.83.

One-Way or Two-Way Operation

The NCHRP research effort for the 1985 HCM found that saturation flow rates decreased by 7.5 percent on one-way streets compared with saturation flow rates on two-way streets. A total of 356 saturation flow vehicles were surveyed on one-way streets in this study. The *Transportation Research Circular 212* technique does not include an adjustment factor for one-way flow on urban streets *(3)*. Moreover, no such adjustment factor is included in the 1985 HCM procedure *(1)*.

The surveys on one-way streets for the research documented in this paper included 1,514 saturation flow vehicles. It was found that the one-way saturation flow rates were 9 percent lower on one-way streets than on comparable two-way streets that had baseline geometric conditions (consistent with the NCHRP surveys). The surveys conducted for the research documented in this paper were conducted on approaches with relatively congested peak-period conditions in major metropolitan areas. Thus, the lower saturation flow rates on one-way streets cannot be attributed to unpressured driving characteristics. Further investigation indicated that lower saturation flow rates could occur on one-way street approaches compared with those on two-way street approaches for two reasons:

- At most urban intersections, there is no difference in side friction between a one-way street approach and a two-way street approach. Many two-way street approaches have raised medians that separate the two directions of flow. Thus, the left lane on a one-way street approach has lateral characteristics comparable to those of the median approach lane on a two-way street approach.
- Most urban one-way streets (including the approaches surveyed) contain relatively frequent traffic signal spacing and good signal progression. Thus, many peak-period motorists leaving an intersection after a red signal phase will adjust their speeds to work into the signal progression band. This frequently results in a slightly reduced departure speed (below the speed that the motorist would otherwise select) and slightly longer headways between vehicles in saturation flow conditions on the approaches.

As a result of the operating conditions stated, an isolated intersection approach along a one-way street may experience lower saturation flow rates compared with those of a comparable two-way street approach; however, route and system efficiency along a one-way street is increased because traffic signal progression reduces the number of stops per vehicle and the overall average vehicle delay. Therefore, consistent with the field data collected during this study and the NCHRP surveys, it is proposed that an adjustment factor of 0.91 (9 percent reduction) be applied to one-way street approach lanes. It appears from observation that applying both the factor for a one-way street and the factor for two or three approach lanes (one-way factor and lane utilization factor) underestimates the critical volumes that can be accommodated. Thus, only the greater of the two factors should be applied in this instance.

Summary of Adjustment Factors

The proposed adjustment factors for each of the 12 characteristics described in this paper are given in the summary in Table 11. They reflect the results of extensive data collection efforts contained in *Transportation Research Circular 212*, the 1985

TABLE 11 SUMMARY OF ADJUSTMENT FACTORS

Adjustment	Range of Values	Factor
1. Lane width	8 to 8.9 ft	0.95
	9 to 9.9 ft	0.98
	13.1 to 15.9 ft	1.05
2. Heavy vehicles	2 percent heavy vehicles	0.98
	6 percent heavy vehicles	0.95
	10 percent heavy vehicles	0.92
3. grade on vertical approach	5 to 6 percent downgrade	0.96
	3 to 4 percent downgrade	0.98
	3 to 4 percent upgrade	0.95
	5 to 6 percent upgrade	0.90
4. Curb Parking		
Approach 1		
0 maneuvers/hr		0.89
20 maneuvers/hr		0.82
40 maneuvers/hr		0.75
Approach 2		
0 maneuvers/hr		0.94
20 maneuvers/hr		0.91
40 maneuvers/hr		0.87
Approach 3		
0 maneuvers/hr		0.97
20 maneuvers/hr		0.95
40 maneuvers/hr		0.93
5. Local bus stop activity		
Approach 1		
0 buses stopping/hr		1.00
20 buses stopping/hr		0.95
40 buses stopping/hr		0.90
Approach 2		
0 buses stopping/hr		1.00
20 buses stopping/hr		0.97
40 buses stopping/hr		0.95
Approach 3		
0 buses stopping/hr		1.00
20 buses stopping/hr		0.99
40 buses stopping/hr		0.97
6. Area type		No adjustment
7. Through-lane utilization		
Two lanes		0.96
Three or more lanes		0.91
8. Turning movements		
Exclusive left-turn lane/protected phase	–	0.97
Exclusive right-turn lane/protected phase	10- to 30-ft radius	0.84
Exclusive right-turn lane/protected phase	30- to 44-ft radius	0.91
Exclusive left-turn lane/permissive phase	–	See Table 3 (times pedestrian factor)
Exclusive right-turn lane/permissive phase	10- to 30-ft radius	0.84 (times pedestrian factor)
Exclusive right-turn lane/permissive phase	30- to 44-ft radius	0.91 (times pedestrian factor)
Optional left-turn through lane/protected phase	–	0.97 (varies by percent turns)
Optional right-turn through lane/protected phase	–	0.84 or 0.91 (varies by percent turns)
Optional left-turn through lane/permissive phase	–	See Table 4 (times pedestrian factor)
Optional right-turn through lane/permissive phase	–	0.84 or 0.91 (varies by percent turns and pedestrian factor)
9. Dual turn lanes (protected phase)/left turns	–	0.94
Dual turn lanes (protected phase)/right turns	–	0.76
10. Pedestrian conflict (permissive left- or right-turn phase)		
100 pedestrians/hr		0.93
200 pedestrians/hr		0.91
400 pedestrians/hr		0.87
800 pedestrians/hr		0.80
1,000 pedestrians/hr		0.76
1,400 pedestrians/hr		0.68
1,800 pedestrians/hr		0.61
2,000 pedestrians/hr		0.57
11. Metropolitan area size	Population of 0 to 20,000	0.83
	Population of 20,000 to 100,000	0.91
12. One-way or two-way operation	One-way operation	0.91

Source: Barton–Aschman Associates, Inc.

HCM, the surveys for the research documented in this paper, and other relevant sources. The proposed factors are similar to the factors contained in the 1985 HCM for curb parking, lane utilization, and exclusive right-turn lanes (protected phase). The factors given in Table 11 are proposed for use in the 1985 HCM technique for signalized intersections and in other signalized capacity analysis techniques in which demand volumes must be adjusted to account for geometric, vehicle characteristic, and environmental conditions.

CONCLUSION

An improved set of saturation flow rates and adjustment factors for signalized intersection capacity analyses has been presented. These values are based on an extensive data base collected throughout the United States. The factors have been structured to be incorporated into calculations of signalized intersection capacity when saturation flow rates must be considered. It is hoped that the results of this research effort will encourage the continuation of inquiry into the values incorporated into signalized intersection capacity analysis procedures.

ACKNOWLEDGMENT

Summarized are the results of a data collection and analysis effort sponsored and conducted by eight offices of Barton–Aschman Associates, Inc., in the United States. The efforts of numerous individuals who directed the field survey and data summary efforts throughout the country are hereby acknowledged. In addition, the suggestions and constructive comments of senior Barton–Aschman staff on the contents of this paper are deeply appreciated. Sincere thanks are due to the management of Barton–Aschman, who underwrote the cost of this effort.

REFERENCES

1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985, pp. 9-1 – 9-84.
2. C. J. Messer and D. B. Fambro. Critical Lane Analysis for Intersection Design. In *Transportation Research Record 644*, TRB, National Research Council, Washington, D.C., 1977, pp. 26–35.
3. *Transportation Research Circular 212: Interim Materials on Highway Capacity.* TRB, National Research Council, Washington, D.C., 1980, pp. 5–36.
4. R. W. Stokes. *Saturation Flows of Exclusive Double Left-Turn Lanes.* Ph.D. dissertation, Texas A&M University, College Station, 1984.
5. K. R. Agent and J. D. Crabtree. *Analysis of Saturation Flow at Signalized Intersections.* Kentucky Transportation Program, University of Kentucky, Lexington, May 1982.
6. JHK & Associates and Northwestern University Traffic Institute. *NCHRP Report: Urban Signalized Intersection Capacity.* TRB, National Research Council, Washington, D.C., 1982.
7. K. R. Agent and J. D. Crabtree. *Analysis of Lost Time at Signalized Intersections.* Kentucky Transportation Research Program, University of Kentucky, Lexington, Feb. 1983.
8. *Special Report 87: Highway Capacity Manual.* HRB, National Research Council, Washington, D.C., 1965, 411 pp.
9. R. L. Carstens. "Some Traffic Parameters at Signalized Intersections." *Traffic Engineering*, Aug. 1971, pp. 33–36.
10. John Hamburg & Associates, Inc. and Price Williams & Associates, Inc. *Micro Assignment Users Guide.* Report URD-JHA-80-2-1. Office of Planning Methods and Support, UMTA, U.S. Department of Transportation, March 1, 1980.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Analysis of the Proposed Use of Delay-Based Levels of Service at Signalized Intersections

DONALD S. BERRY AND RONALD C. PFEFER

Stopped delay calculations were made for many combinations of cycle length, ratio of effective green time to cycle length (*g/C* ratio), and quality of signal progression, to aid in the following: identifying the relationships between computed delay and volume-to-capacity (*v/c*) ratio for lane groups at signalized intersections; determining how to solve for *v/c* ratio and service flow rate when the desired level-of-service (LOS) delay value is known; examining methods for using the delay-based LOSs in intersection design, both geometrics and signal timing. Sensitivity analyses reveal that many combinations of long cycle lengths, low *g/C* ratios, and adverse progression result in such high delays that LOS levels of A, B, and C are unattainable, even at low *v/c* ratios. Computer-generated tabulations are proposed to aid in solving for *v/c* ratios and maximum flow rates associated with desired delay-based LOSs.

In the new Highway Capacity Manual (HCM), use of computed average stopped delay for determining the level of service (LOS) at signalized intersections is specified *(1)*. In the process of the refinement of the methodology for inclusion in the new HCM, several issues were raised. Among these were the following:

- What role, if any, should the volume-to-capacity (*v/c*) ratio have in the process?
- What are the appropriate average delay ranges to use in defining LOS?
- How readily can the method be employed in intersection design when the analyst seeks a direct indication of the maximum service volume flow rate for a given condition and stated LOS?

For the purpose of examining these issues, a preliminary investigation of the LOS module is described in this paper. First, basic tabulations developed for this analysis are introduced and explained. Second, the results are presented as a series of sensitivity analyses, with significant attributes highlighted. This is followed by a series of observations on the results. The paper concludes with some suggestions on design applications.

BASIC TABULATIONS

Basic tabulations were prepared by computing stopped delay for different combinations of cycle length, ratios of effective green time to cycle length (*g/C* ratios), and categories of signal progression. Eleven combinations of cycle length and *g/C* ratios were used, as follows:

Cycle Length (sec)	*g/C Ratio*
60	0.3, 0.5, 0.6, 0.7
90	0.2, 0.5, 0.7
120	0.2, 0.4, 0.5, 0.6

The calculations were made for a typical intersection approach with two through lanes, saturation flow of 3,200 vehicles per hour of green time (vphg), and pretimed signal control.

Stopped delay for random arrivals was calculated by using Equation 9-18 of the HCM. The results were then multiplied by adjustment factors for quality of signal progression, as given in Table 9-13 of the HCM. Equation 9-18 is as follows *(1)*:

$$d = 0.38C\ \{(1 - g/C)^2/[1 - g/C(x)]\} + 173x^2\ \{(x - 1) + [(x - 1)^2 + (16x/c)]^{1/2}\}$$

where

- d = average stopped delay for a lane group (sec/veh) for random arrivals;
- C = cycle length (sec);
- g/C = green ratio for the lane group; the ratio of effective green to cycle length;
- x = v/c ratio for the lane group;
- v = flow rate (vph); and
- c = capacity of the lane group (vph).

Progression adjustment factors are given in Table 9-13 of the HCM for five qualities of progression (QP1 to QP5). These factors vary from 1.85 for very adverse progression (QP1) to 0.40 for excellent progression (QP5). These adjustment factors also vary somewhat with *v/c* ratio and by type of traffic control signal, as indicated by the data in Table 9-13 of the HCM.

The tables in Appendix A of this paper were created by using an IBM personal computer and the Lotus 1-2-3 spreadsheet. The spreadsheet was created by using the table lookup functions that are a part of the Lotus 1-2-3 package. The initial tabulation (given in the tops of Tables A-1 to A-11) of delays for various combinations of *v/c* and quality of progression was developed, as noted previously, by using Equation 9-18 of the HCM *(1)*. The tabulations of maximum service volume and *v/c* for various levels of delay and quality of progression (given in the bottoms of Tables A-1 to A-11) were created by looking up

D. S. Berry, Civil Engineering, Northwestern University, 2520 Park Pl., Evanston, Ill. 60201. R. C. Pfefer, Research and Development, The Traffic Institute, Northwestern University, 405 Church St., Evanston, Ill. 60201.

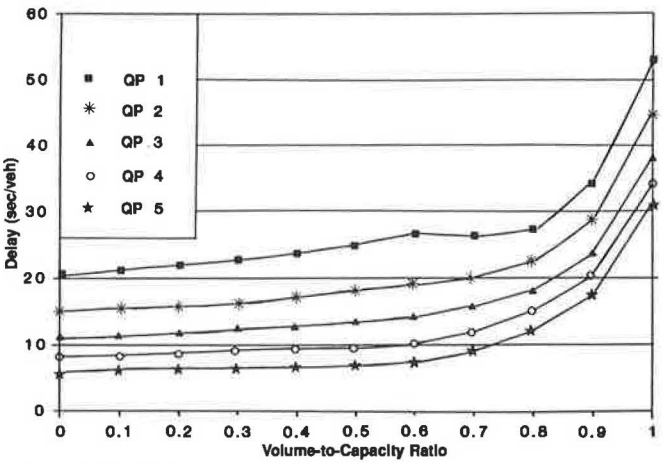


FIGURE 1 Delay versus *v/c* ratio [*c* = 60 sec, *g/C* = 0.3, and *s* = 3,200 vph (two lanes)].

the values in the initial tabulation and interpolating to arrive at the exact figures.

Two tables were created for each combination of cycle length and *v/c* ratio and placed on a single sheet, as indicated by the data in Table A-1 of Appendix A. The top table provides a tabulation of average stopped delay, for each quality of progression (1 through 5) for *v/c* ratios between 0 and 1 in increments of 0.1. The table includes a listing of both the uniform delay (U.D.) and random delay (R.D.) elements of Equation 9-18, as well as the flow rate associated with the *v/c* ratio. The lower table is a tabulation of maximum service volume *(sv)* flow rates and *v/c* ratios up to 1.0 associated with 5-sec increments of delay from 5 to 60 sec per vehicle. Columns are included for each quality of progression. Blank cells in the tables result either for delay levels that are unattainable, or for which *v/c* is greater than 1.0.

Basic attributes of the delay equation are evident, as shown graphically in Figure 1. There is a minimum level of average stopped delay associated with the cycle length, quality of progression, and *g/C* ratio, below which one cannot go as volume approaches zero. As a result, the *v/c* ratio has little influence on average stopped delay at *v/c* levels below 0.6 or 0.7. It is at levels above this that delay increases rapidly. Delays at *v/c* values of 1.10 are much higher than delay at *v/c* of 1.0, as indicated by the data in Table A-4.

Figure 1 shows that the effect of quality of progression on average stopped delay can be significant. The extrapolation of progression adjustment factors that was made in creating Table 9-13 of the HCM highlights a small area of concern in some areas of computation. There are some cases in which the decrease in the progression adjustment factor that results when *v/c* increases above 0.6 more than offsets the associated increase in delay. As a result, computed delay may decrease as *v/c* increases above 0.60. A case in point is the tabulation for 120-sec cycle length at *g/C* of 0.4 with Quality of Progression 1 (QP1). Computed average stopped delays for *v/c* values of 0.7 and 0.8 are lower than those computed on either side. This discontinuity is not likely to be a problem at other levels of quality of progression because the increments for the other progression adjustment factors are much less than for QP1. Some modification of Table 9-13 of the HCM is needed *(1)*.

SENSITIVITY ANALYSIS (SATURATION FLOW = 3,200 VPHG)

Table 1 gives, for three cycle lengths and several *g/C* ratios, the *v/c* ratios associated with each of nine delay values as com-

TABLE 1 EFFECTS OF CYCLE LENGTH AND *g/C* ON *v/C* VALUES FOR AVERAGE STOPPED DELAY

CYCLE		CAPACITY	AVERAGE STOPPED DELAY, SEC/VEH								
LENGTH	G/C	VPH	5	10	15	20	25	30	35	40	45
60 Sec	.70	2240	.72	.90	.95	.99	>1	>1	>1	>1	>1
	.60	1920	.40	.83	.92	.96	>1	>1	>1	>1	>1
	.50	1600	-	.71	.88	.93	.97	>1	>1	>1	>1
	.30	960	-	-	.65	.84	.91	.94	.98	>1	>1
90 Sec	.70	2240	.51	.84	.92	.96	>1	>1	>1	>1	>1
	.50	1600	-	.28	.74	.88	.93	.97	>1	>1	>1
	.20	640	-	-	-	-	.51	.77	.86	.91	.94
120 Sec	.60	1920	-	.43	.77	.89	.93	.97	>1	>1	>1
	.50	1600	-	-	.46	.76	.88	.93	.96	1.0	>1
	.40	1280	-	-	-	.43	.74	.86	.92	.95	.99
	.20	640	-	-	-	-	-	.13	.63	.80	.87

Note: For Quality of Progression 3 (random approach) using adjustment factors from Table 9-13 *(1)* and saturation flow = 3,200 vphg.

puted for Quality of Progression 3 (QP3). These data were obtained from the tables in Appendix A. It is apparent from the data in Table 1 that short cycle lengths and high *g/C* ratios are associated with low delays, even at high *v/c* ratios. For this type of signal timing, an LOS A delay range of 0 to 5 sec/veh is suitable and was adopted for Table 9-1 of the HCM. The data in Table 1 also indicate that, conversely, some long cycle lengths and low *g/C* ratios result in such high delays that the lower delay levels associated with certain high levels of service such as LOS A, B, and C are not attainable. This could influence use of shorter cycle lengths.

Tables 2, 3, and 4 present additional results for each of five qualities of signal progression for different combinations of cycle length and *g/C* ratios. It is apparent that quality of progression is a major factor in affecting *v/c* ratios and the resulting *sv*'s that can be accommodated at each LOS delay range.

Table 4 gives a summary of the effects of using longer cycle lengths and shorter *g/C* ratios such as those used in multiphase signal timing. The data in this table indicate that with a timing plan with a 120-sec cycle, *g/C* of 0.20, and random arrivals (QP3), the *v/c* value must not exceed 0.13 in order to keep the stopped delay below 30 sec/veh. However, if *g/C* is increased to 0.40, with the same 120-sec cycle and QP3, it is possible to attain a *v/c* ratio of 0.86 and still not exceed a delay of 30 sec/veh.

In Table 5, the use of data from Tables 2 and 3 in comparing *v/c* values and *sv*'s for different quality of progression values is explored further. In Case 1, a change from QP1 to QP5 results in a large increase in *sv* for an LOS A delay value of 5.0 sec/veh (from 739 to 1,859 vph). In Case 3, (90-sec cycle and *g/C* of 0.50), a change from QP3 to QP5 results in only a moderate increase in *sv* (from 1,184 to 1,424) when using the LOS B delay criterion of 15 sec/veh.

OBSERVATIONS

Role of *v/c*

Some continue to suggest that *v/c* be used in lieu of average stopped delay as the basis for LOS. The Transportation Research Board Committee on Highway Capacity and Quality of Service continues to stand firm in its use of stopped delay as the measure for LOS. The analysis in this paper demonstrates one value of using the delay measure. Long cycle lengths and low *g/C* values have relatively high levels of delay associated with relatively low *v/c* values. Use of *v/c* for defining LOS could easily result in excessive and costly delays, which are already a matter for concern in some jurisdictions currently using long cycle lengths. The use of average delay as the LOS

TABLE 2 EFFECTS OF *g/C* ON *v/C* VALUES BY LEVEL OF DELAY FOR 60-SEC CYCLE LENGTH AND SATURATION FLOW = 3,200 VPHG

G/C	QUALITY OF PROGRESS	AVERAGE STOPPED DELAY, SEC/VEH								
		5	10	15	20	25	30	35	40	45
.70	1	.33	.82	.91	.94	.97	1.0	>1	>1	>1
	2	.57	.86	.93	.96	1.0	>1	>1	>1	>1
c=2240	3	.72	.90	.95	.99	>1	>1	>1	>1	>1
	4	.79	.92	.96	>1	>1	>1	>1	>1	>1
	5	.83	.93	.98	>1	>1	>1	>1	>1	>1
.60	1	-	.50	.84	.91	.94	.97	>1	>1	>1
	2	-	.76	.90	.93	.98	>1	>1	>1	>1
c=1920	3	.40	.83	.92	.96	>1	>1	>1	>1	>1
	4	.64	.88	.94	.98	>1	>1	>1	>1	>1
	5	.75	.91	.95	1.0	>1	>1	>1	>1	>1
.50	1	-	-	.54	.85	.91	.94	.97	>1	>1
	2	-	.43	.82	.91	.94	.97	>1	>1	>1
c=1600	3	-	.71	.88	.93	.97	>1	>1	>1	>1
	4	.34	.81	.91	.97	>1	>1	>1	>1	>1
	5	.63	.86	.93	>1	>1	>1	>1	>1	>1
.30	1	-	-	-	-	.50	.84	.91	.93	.96
	2	-	-	-	.68	.85	.91	.94	.97	1.0
c=960	3	-	-	.65	.84	.91	.94	.98	>1	>1
	4	-	.54	.80	.90	.93	.97	>1	>1	>1
	5	-	.72	.85	.92	.95	.99	>1	>1	>1

TABLE 3 EFFECTS OF *g/C* ON *v/C* VALUES BY LEVEL OF DELAY FOR 90-SEC CYCLE LENGTH AND SATURATION FLOW = 3,200 VPHG

G/C	QUALITY OF PROGRESS	AVERAGE STOPPED DELAY, SEC/VEH								
		5	10	15	20	25	30	35	40	45
.70	1	-	.57	.84	.91	.94	.97	>1	>1	>1
	2	.24	.78	.90	.94	.97	>1	>1	>1	>1
c=2240	3	.51	.84	.92	.96	>1	>1	>1	>1	>1
	4	.68	.88	.94	.98	>1	>1	>1	>1	>1
	5	.76	.91	.95	1.0	>1	>1	>1	>1	>1
.50	1	-	-	-	.40	.81	.90	.93	.95	.98
	2	-	-	.44	.80	.90	.93	.96	1.0	>1
c=1600	3	-	.28	.74	.88	.93	.97	>1	>1	>1
	4	-	.65	.84	.92	.95	.99	>1	>1	>1
	5	.19	.76	.89	.93	.97	>1	>1	>1	>1
.20	1	-	-	-	-	-	-	-	-	.44
	2	-	-	-	-	-	.07	.59	.83	.89
c=640	3	-	-	-	-	.51	.77	.86	.91	.94
	4	-	-	-	.65	.79	.86	.91	.94	.97
	5	-	-	.64	.78	.85	.91	.94	.97	>1

TABLE 4 EFFECTS OF *g/C* ON *v/C* VALUES BY LEVEL OF DELAY FOR 120-SEC CYCLE LENGTH AND SATURATION FLOW = 3,200 VPHG

G/C	QUALITY OF PROGRESS	AVERAGE STOPPED DELAY, SEC/VEH								
		5	10	15	20	25	30	35	40	45
.60	1	-	-	.16	.52	.82	.90	.93	.96	.99
	2	-	.02	.54	.81	.90	.94	.97	1.0	>1
c=1920	3	-	.43	.77	.89	.93	.97	>1	>1	>1
	4	-	.68	.84	.92	.96	1.0	>1	>1	>1
	5	.37	.78	.89	.94	.98	>1	>1	>1	>1
.50	1	-	-	-	-	.31	.56	.85	.91	.94
	2	-	-	-	.44	.76	.88	.92	.95	.99
c=1600	3	-	-	.46	.76	.88	.93	.96	1.0	>1
	4	-	.35	.72	.85	.92	.96	.99	>1	>1
	5	-	.65	.82	.90	.94	.98	>1	>1	>1
.40	1	-	-	-	-	-	-	.32	.56	.88
	2	-	-	-	-	.28	.61	.84	.91	.94
c=1280	3	-	-	-	.43	.74	.86	.92	.95	.99
	4	-	-	.50	.74	.85	.92	.95	.99	>1
	5	-	.32	.72	.83	.91	.94	.98	>1	>1
.20	1	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	.07	.53
c=640	3	-	-	-	-	-	.13	.63	.80	.87
	4	-	-	-	-	.61	.74	.83	.90	.93
	5	-	-	-	.65	.77	.84	.90	.93	.96

TABLE 5 EXAMPLES: EFFECTS OF IMPROVING QUALITY OF PROGRESSION ON v/C RATIOS AND SERVICE VOLUMES

CASE NO.	CYCLE LENGTH	G/C	AVERAGE STOPPED DELAY, Sec/veh	QUALITY OF PROGRESSION	v/c	SERVICE VOLUME
1	60	.70	5	QP_1	.33	739
				QP_5	.83	1859
2	90	.50	10	QP_3	.28	448
				QP_5	.76	1217
3	90	.50	15	QP_3	.74	1184
				QP_5	.89	1424

Note: The source of the data is Tables 2 and 3.

criterion should encourage selection of the shortest cycle lengths possible.

However, given the sensitivity of delay to high *v/c* ratios, the designer should be discouraged from designing to relatively high *v/c* ratios even though they can be associated with delays approximating LOS C with the criteria currently proposed. This would be in order to create a supplementary factor of safety. To accomplish this, it may be desirable to add guidelines that specify *v/c* ratios as a supplement to the delay criteria for LOS in the design range. Suggested for trial are the following:

LOS	*v/c Not to Exceed*
A	0.85
B	0.90
C	0.93

Using such supplementary criteria will avoid the situation in which a designer, needing to service high demands, will select high *v/c* values that are possible at some LOS delay ranges. Use of such high *v/c* ratios could readily result in excessive delays when small increases in volume occur that exceed assumed design volumes.

Delay Criteria for LOS

The tabulations in this paper have already been useful to the Committee on Highway Capacity and Quality of Service in finalizing the delay criteria used for LOS in Table 9-1 of Chapter 9 of the HCM, Level-of-Service Criteria for Signalized Intersections.

Design Applications

There are a variety of situations for which analyses of a signalized intersection may take place. The majority may be classified into three types:

- Operational analysis,
- Intersection design or redesign, and
- Signal control design or redesign.

An operational analysis has as its objective the determination of LOS for a stated set of geometric, traffic, and signal timing conditions. The intersection design or redesign classification involves arriving at an optimum set of geometric and control design parameters for a given set of traffic conditions and desired LOS. The signal control design or redesign classification is similar to the preceding one, except that geometric conditions are essentially taken as a given and traffic signal control parameters are determined for a desired LOS.

In addition to the three major cases noted, there is present within each class the implicit objective of finding the maximum *v/c* and *sv* associated with a specific LOS.

The steps involved in applying the proposed methodology to determine maximum *v/c* and *sv*'s associated with a desired LOS delay value are shown in Figures 9-14*a*, 9-14*b*, and 9-14*c* of the HCM. Figure 2 in this paper (identical to Figure 9-14*a* of HCM) shows the steps to follow when the geometric configuration, the signal timing, and the desired LOS delay value are known, and the *v/c* ratio and maximum *sv*'s are the unknowns. In this case, the delay tabulation is quickly generated, as presented in the upper part of Table A-1. The lower part of Table A-1 gives the *v/c* and *sv* values associated with desired LOS delay values.

Sample calculation 6 of the HCM is an example of use of the computer in generating data for Step 6 of Figure 9-14*a*. Inputs are geometrics (two through lanes), cycle length (90 sec), *g/C* ratio (0.50), and a desired LOS of B (maximum delay of 15 sec/veh). The computer generates the output given in Table A-6 (also shown in the HCM as Figure 9-36). The data in the lower part of the table indicates that for a LOS delay of 15.0 sec/veh and QP3 (random arrivals), the maximum *v/c* ratio is 0.74 and the maximum service flow rate is 1,183 vph.

When the geometrics or the signal timing are the unknown elements, as in Figures 9-14*b* and 9-14*c* of the HCM, alternative procedures are needed. The generation of tables by computer permits the analyst to quickly identify the ranges of geometric and/or signal design that will accommodate the given *sv*'s at the desired LOS delay value.

Tables of this type may be prepared in a variety of formats. The appropriate one is to be determined in the context of the overall automated package. The ability to generate a tabulation of this type facilitates the performance of sensitivity analyses by the analyst. In the context of a comprehensive applications package, this type of computation might be most effectively employed in a built-in optimization process in which the computation for alternative phasing arrangements might be performed by the computer, which then prints out the results, identifying an optimum configuration according to a specified objective function. This would further reduce the number of data entry steps required of the analyst. The analyst would retain the flexibility to choose from the nonoptimum alternatives if constraints are present that dictate such an action.

Using delay as the criterion for level of service will undoubtedly affect procedures for selecting cycle length and the phasing for isolated signals. Tables of the type displayed in this paper may be helpful in preliminary analyses of alternative designs and signal timing plans. Because short-cycle, two-phase signals produce the lowest delays, there may be a shift toward use of two-phase signals and toward alternative methods for providing for left turns. There also may be more

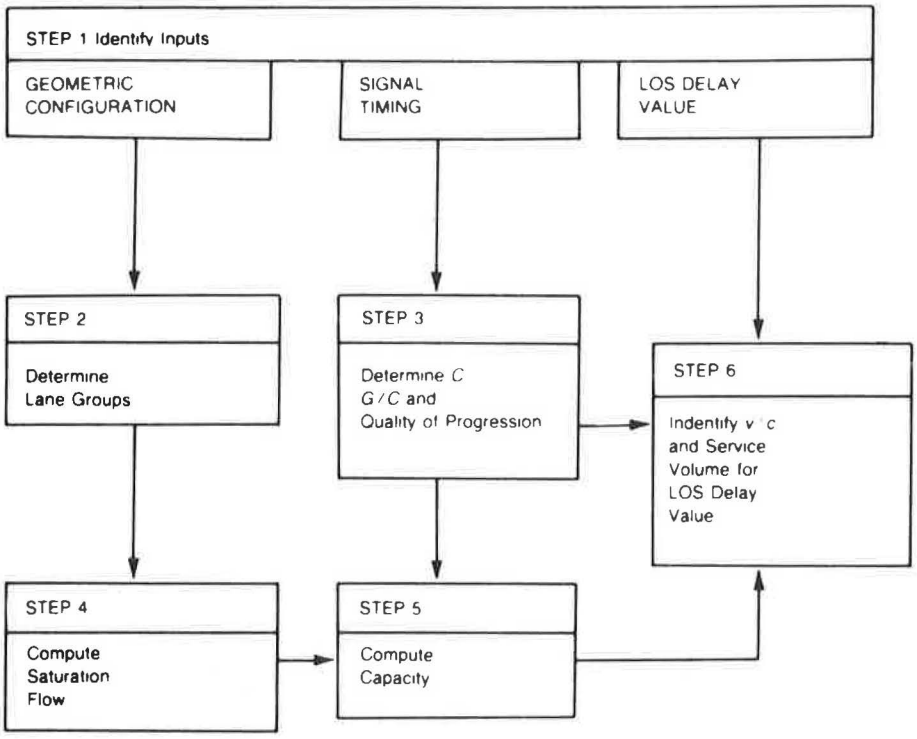


FIGURE 2 Determining *v/c* ratios and service flow rates.

effort devoted to attaining a higher quality of progression in order to increase service volumes and reduce delay.

REFERENCE

1. Signalized Intersections. In *Special Report 209: Highway Capacity Manual*, TRB, National Research Council, Washington, D.C., 1985, Chapter 9.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

APPENDIX A

Basic Tabulations

Table	Cycle Length	g/C
A-1	60	0.3
A-2	60	0.5
A-3	60	0.6
A-4	60	0.7
A-5	90	0.2
A-6	90	0.5
A-7	90	0.7
A-8	120	0.2
A-9	120	0.4
A-10	120	0.5
A-11	120	0.6

AVERAGE STOPPED DELAY FOR:

TABLE A-1

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 60
G/C= 0.3
CAPACITY 960

				QUALITY OF PROGRESSION				
V/C	U.D.	R.D.	FLOW RATE	1	2	3	4	5
0	11.17	0.00	0	20.67	15.08	11.17	8.04	5.92
0.1	11.52	.00	96	21.31	15.55	11.52	8.29	6.11
0.2	11.89	0.01	192	22.01	16.06	11.90	8.57	6.31
0.3	12.28	0.06	288	22.81	16.65	12.33	8.88	6.54
0.4	12.70	0.15	384	23.77	17.35	12.85	9.25	6.81
0.5	13.14	0.36	480	24.98	18.23	13.50	9.72	7.16
0.6	13.62	0.77	576	26.62	19.43	14.39	10.36	7.63
0.7	14.14	1.60	672	26.44	20.15	15.74	12.12	9.44
0.8	14.70	3.43	768	27.19	22.11	18.13	14.86	12.14
0.9	15.30	8.14	864	34.00	28.14	23.45	20.16	17.59
1	15.96	22.33	960	53.61	45.19	38.29	34.46	31.40
1.1	16.67	56.17	1056	101.98	85.95	72.84	65.56	59.73

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

	MAXIMUM STOPPED	QUALITY OF PROGRESSION											
		QP=1		QP=2		QP=3		QP=4		QP=5			
LOS	DELAY	MAX SV	V/C	*MAX SV	V/C	*MAX SV	V/C	*MAX SV	V/C	*MAX SV	V/C		
A	5												
	10							522	0.54	692	0.72		
B	15					619	0.65	770	0.80	818	0.85		
	20			652	0.68	802	0.84	861	0.90	881	0.92		
C	25	481	0.50	814	0.85	874	0.91	896	0.93	916	0.95		
	30	808	0.84	874	0.91	906	0.94	930	0.97	950	0.99		
	35	869	0.91	903	0.94	939	0.98						
D	40	893	0.93	931	0.97								
	45	918	0.96	959	1.00								
	50	942	0.98										
	55												
E	60												

AVERAGE STOPPED DELAY FOR:

TABLE A-2

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 60
G/C= 0.5
CAPACITY 1600

V/C	U.D.	R.D.	FLOW RATE	QUALITY OF PROGRESSION 1	2	3	4	5
0	5.70	0.00	0	10.55	7.70	5.70	4.10	3.02
0.1	6.00	.00	160	11.10	8.10	6.00	4.32	3.18
0.2	6.33	0.01	320	11.73	8.56	6.34	4.57	3.36
0.3	6.71	0.03	480	12.47	9.10	6.74	4.85	3.57
0.4	7.13	0.09	640	13.35	9.74	7.22	5.20	3.83
0.5	7.60	0.22	800	14.46	10.55	7.82	5.63	4.14
0.6	8.14	0.46	960	15.92	11.62	8.61	6.20	4.56
0.7	8.77	0.97	1120	16.36	12.47	9.74	7.50	5.84
0.8	9.50	2.11	1280	17.42	14.17	11.61	9.52	7.78
0.9	10.36	5.30	1440	22.72	18.80	15.67	13.47	11.75
1	11.40	17.30	1600	40.18	33.87	28.70	25.83	23.53
1.1	12.67	51.27	1760	89.51	75.44	63.93	57.54	52.43

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

LOS	MAXIMUM STOPPED DELAY	QP=1 MAX SV	V/C	QP=2 MAX SV	V/C	QP=3 MAX SV	V/C	QP=4 MAX SV	V/C	QP=5 MAX SV	V/C
A	5							549	0.34	1015	0.63
	10			691	0.43	1142	0.71	1299	0.81	1369	0.86
B	15	859	0.54	1309	0.82	1414	0.88	1460	0.91	1484	0.93
	20	1358	0.85	1453	0.91	1493	0.93	1525	0.95	1552	0.97
C	25	1461	0.91	1506	0.94	1555	0.97	1589	0.99		
	30	1507	0.94	1559	0.97						
	35	1553	0.97								
D	40	1598	1.00								
	45										
	50										
	55										
E	60										

AVERAGE STOPPED DELAY FOR:

TABLE A-4

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 60
G/C= 0.7
CAPACITY 2240

V/C	U.D.	R.D.	FLOW RATE	QUALITY OF PROGRESSION 1	2	3	4	5
0	2.05	0.00	0	3.80	2.77	2.05	1.48	1.09
0.1	2.21	.00	224	4.08	2.98	2.21	1.59	1.17
0.2	2.39	0.01	448	4.43	3.23	2.39	1.72	1.27
0.3	2.60	0.02	672	4.85	3.54	2.62	1.89	1.39
0.4	2.85	0.07	896	5.39	3.94	2.92	2.10	1.55
0.5	3.16	0.15	1120	6.13	4.47	3.31	2.38	1.75
0.6	3.54	0.33	1344	7.16	5.22	3.87	2.79	2.05
0.7	4.02	0.70	1568	7.93	6.04	4.72	3.63	2.83
0.8	4.66	1.53	1792	9.29	7.55	6.19	5.08	4.15
0.9	5.55	3.95	2016	13.77	11.39	9.49	8.16	7.12
1	6.84	14.62	2240	30.05	25.32	21.46	19.32	17.60
1.1	8.92	48.91	2464	80.96	68.24	57.83	52.04	47.42

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

LOS	MAXIMUM STOPPED DELAY	QP=1 MAX SV	V/C	QP=2 MAX SV	V/C	QP=3 MAX SV	V/C	QP=4 MAX SV	V/C	QP=5 MAX SV	V/C
A	5	734	0.33	1278	0.57	1611	0.72	1780	0.79	1856	0.83
	10	1828	0.82	1935	0.86	2025	0.90	2053	0.92	2078	0.93
B	15	2033	0.91	2074	0.93	2119	0.95	2153	0.96	2184	0.98
	20	2102	0.94	2154	0.96	2213	0.99				
C	25	2171	0.97	2235	1.00						
	30	2239	1.00								
	35										
D	40										
	45										
	50										
	55										
E	60										

AVERAGE STOPPED DELAY FOR:

TABLE A-3

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 60
G/C= 0.6
CAPACITY 1920

V/C	U.D.	R.D.	FLOW RATE	QUALITY OF PROGRESSION 1	2	3	4	5
0	3.65	0.00	0	6.75	4.92	3.65	2.63	1.93
0.1	3.88	.00	192	7.18	5.24	3.88	2.79	2.06
0.2	4.15	0.01	384	7.68	5.61	4.15	2.99	2.20
0.3	4.45	0.03	576	8.28	6.04	4.48	3.22	2.37
0.4	4.80	0.08	768	9.02	6.58	4.88	3.51	2.58
0.5	5.21	0.18	960	9.97	7.28	5.39	3.88	2.86
0.6	5.70	0.39	1152	11.26	8.22	6.09	4.38	3.23
0.7	6.29	0.81	1344	11.93	9.09	7.10	5.47	4.26
0.8	7.02	1.77	1536	13.18	10.72	8.79	7.21	5.89
0.9	7.93	4.52	1728	18.06	14.95	12.45	10.71	9.34
1	9.12	15.79	1920	34.88	29.40	24.91	22.42	20.43
1.1	10.73	49.91	2112	84.90	71.56	60.64	54.58	49.73

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

LOS	MAXIMUM STOPPED DELAY	QP=1 MAX SV	V/C	QP=2 MAX SV	V/C	QP=3 MAX SV	V/C	QP=4 MAX SV	V/C	QP=5 MAX SV	V/C
A	5			46	0.02	814	0.42	1261	0.66	1431	0.75
	10	964	0.50	1451	0.76	1599	0.83	1689	0.88	1739	0.91
B	15	1608	0.84	1729	0.90	1767	0.92	1798	0.94	1826	0.95
	20	1750	0.91	1795	0.93	1844	0.96	1880	0.98	1913	1.00
C	25	1807	0.94	1962	0.97						
	30	1864	0.97								
	35										
D	40										
	45										
	50										
	55										
E	60										

AVERAGE STOPPED DELAY FOR:

TABLE A-5

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 90
G/C= 0.2
CAPACITY 640

V/C	U.D.	R.D.	FLOW RATE	QUALITY OF PROGRESSION 1	2	3	4	5
0	21.89	0.00	0	40.49	29.55	21.89	15.76	11.60
0.1	22.33	.00	64	41.32	30.16	22.34	16.08	11.84
0.2	22.80	0.02	128	42.22	30.81	22.82	16.43	12.10
0.3	23.29	0.08	192	43.23	31.55	23.37	16.83	12.39
0.4	23.79	0.23	256	44.44	32.43	24.02	17.29	12.73
0.5	24.32	0.53	320	45.98	33.55	24.85	17.89	13.17
0.6	24.87	1.14	384	48.13	35.12	26.01	18.73	13.79
0.7	25.45	2.36	448	46.73	35.60	27.81	21.42	16.69
0.8	26.06	4.98	512	46.55	37.86	31.03	25.45	20.79
0.9	26.69	11.25	576	55.02	45.53	37.94	32.63	28.46
1	27.36	27.35	640	76.60	64.56	54.71	49.24	44.87
1.1	28.06	61.47	704	125.34	105.65	89.53	80.58	73.42

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

LOS	MAXIMUM STOPPED DELAY	QP=1 MAX SV	V/C	QP=2 MAX SV	V/C	QP=3 MAX SV	V/C	QP=4 MAX SV	V/C	QP=5 MAX SV	V/C
A	5										
	10										
B	15									411	0.64
	20							414	0.65	500	0.78
C	25					328	0.51	505	0.79	547	0.85
	30			48	0.07	491	0.77	553	0.86	582	0.91
	35			379	0.59	549	0.86	585	0.91	602	0.94
D	40			530	0.83	584	0.91	604	0.94	621	0.97
	45	279	0.44	572	0.89	603	0.94	624	0.97		
	50	538	0.84	591	0.92	622	0.97				
	55	576	0.90	608	0.95						
E	60	591	0.92	625	0.98						

AVERAGE STOPPED DELAY FOR:

TABLE A-6

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 90
G/C= 0.5
CAPACITY 1600

V/C	U.D.	R.D.	FLOW RATE	QUALITY OF PROGRESSION 1	2	3	4	5
0	8.55	0.00	0	15.82	11.54	8.55	6.16	4.53
0.1	9.00	.00	160	16.65	12.15	9.00	6.48	4.77
0.2	9.50	0.01	320	17.59	12.84	9.51	6.85	5.04
0.3	10.06	0.03	480	18.67	13.62	10.09	7.27	5.35
0.4	10.69	0.09	640	19.94	14.55	10.78	7.76	5.71
0.5	11.40	0.22	800	21.49	15.68	11.62	8.36	6.16
0.6	12.21	0.46	960	23.45	17.11	12.68	9.13	6.72
0.7	13.15	0.97	1120	23.73	18.08	14.12	10.88	8.47
0.8	14.25	2.11	1280	24.55	19.96	16.36	13.42	10.96
0.9	15.55	5.30	1440	30.23	25.02	20.85	17.93	15.64
1	17.10	17.30	1600	48.16	40.59	34.40	30.96	28.21
1.1	19.00	51.27	1760	98.37	82.92	70.27	63.24	57.62

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

LOS	MAXIMUM STOPPED DELAY	QP=1 MAX SV	QP=1 V/C	QP=2 MAX SV	QP=2 V/C	QP=3 MAX SV	QP=3 V/C	QP=4 MAX SV	QP=4 V/C	QP=5 MAX SV	QP=5 V/C
A	5									296	0.19
	10					455	0.28	1040	0.65	1218	0.76
B	15			703	0.44	1183	0.74	1336	0.84	1418	0.89
	20	646	0.40	1281	0.80	1410	0.88	1465	0.92	1496	0.93
C	25	1293	0.81	1439	0.90	1489	0.93	1527	0.95	1559	0.97
	30	1434	0.90	1491	0.93	1548	0.97	1588	0.99		
	35	1483	0.93	1543	0.96						
D	40	1527	0.95	1594	1.00						
	45	1572	0.98								
	50										
	55										
E	60										

AVERAGE STOPPED DELAY FOR:

TABLE A-7

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 90
G/C= 0.7
CAPACITY= 2240

V/C	U.D.	R.D.	FLOW RATE	QUALITY OF PROGRESSION 1	2	3	4	5
0	3.08	0.00	0	5.69	4.16	3.08	2.22	1.63
0.1	3.31	.00	224	6.12	4.47	3.31	2.38	1.75
0.2	3.58	0.01	448	6.63	4.84	3.59	2.58	1.90
0.3	3.90	0.02	672	7.25	5.29	3.92	2.82	2.08
0.4	4.28	0.07	896	8.03	5.86	4.34	3.13	2.30
0.5	4.74	0.15	1120	9.05	6.60	4.89	3.52	2.59
0.6	5.31	0.33	1344	10.43	7.61	5.64	4.06	2.99
0.7	6.04	0.70	1568	11.31	8.62	6.73	5.18	4.04
0.8	7.00	1.53	1792	12.79	10.40	8.52	6.99	5.71
0.9	8.32	3.95	2016	17.79	14.72	12.27	10.55	9.20
1	10.26	14.62	2240	34.83	29.36	24.88	22.39	20.40
1.1	13.38	48.91	2464	87.20	73.50	62.29	56.06	51.08

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

LOS	MAXIMUM STOPPED DELAY	QP=1 MAX SV	QP=1 V/C	QP=2 MAX SV	QP=2 V/C	QP=3 MAX SV	QP=3 V/C	QP=4 MAX SV	QP=4 V/C	QP=5 MAX SV	QP=5 V/C
A	5			527	0.24	1153	0.51	1531	0.68	1697	0.76
	10	1274	0.57	1742	0.78	1880	0.84	1981	0.88	2032	0.91
B	15	1891	0.84	2020	0.90	2065	0.92	2100	0.94	2132	0.95
	20	2045	0.91	2097	0.94	2153	0.96	2195	0.98	2232	1.00
C	25	2111	0.94	2173	0.97						
	30	2176	0.97								
	35										
D	40										
	45										
	50										
	55										
E	60										

AVERAGE STOPPED DELAY FOR:

TABLE A-8

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 120
G/C= 0.2
CAPACITY 640

V/C	U.D.	R.D.	FLOW RATE	QUALITY OF PROGRESSION 1	2	3	4	5
0	29.18	0.00	0	53.99	39.40	29.18	21.01	15.47
0.1	29.78	.00	64	55.10	40.21	29.78	21.44	15.78
0.2	30.40	0.02	128	56.28	41.07	30.42	21.90	16.12
0.3	31.05	0.08	192	57.59	42.03	31.13	22.41	16.50
0.4	31.72	0.23	256	59.11	43.13	31.95	23.00	16.93
0.5	32.43	0.53	320	60.98	44.50	32.96	23.73	17.47
0.6	33.16	1.14	384	63.46	46.31	34.31	24.70	18.18
0.7	33.93	2.36	448	60.98	46.46	36.30	27.95	21.78
0.8	34.74	4.98	512	59.58	48.46	39.72	32.57	26.61
0.9	35.59	11.25	576	67.92	56.21	46.84	40.28	35.13
1	36.48	27.35	640	89.37	75.32	63.83	57.45	52.34
1.1	37.42	61.47	704	138.44	116.68	98.88	89.00	81.09

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

LOS	MAXIMUM STOPPED DELAY	QP=1 MAX SV	QP=1 V/C	QP=2 MAX SV	QP=2 V/C	QP=3 MAX SV	QP=3 V/C	QP=4 MAX SV	QP=4 V/C	QP=5 MAX SV	QP=5 V/C
A	5										
	10										
B	15										
	20									416	0.65
C	25							390	0.61	491	0.77
	30					86	0.13	476	0.74	537	0.84
	35					406	0.63	532	0.83	575	0.90
D	40			48	0.07	515	0.80	574	0.90	594	0.93
	45			338	0.53	559	0.87	594	0.93	613	0.96
	50			525	0.82	588	0.92	612	0.96	631	0.99
	55	58	0.09	566	0.88	607	0.95	631	0.99		
E	60	287	0.45	589	0.92	626	0.98				

AVERAGE STOPPED DELAY FOR:

TABLE A-9

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 120
G/C= 0.4
CAPACITY 1280

V/C	U.D.	R.D.	FLOW RATE	QUALITY OF PROGRESSION 1	2	3	4	5
0	16.42	0.00	0	30.37	22.16	16.42	11.82	8.70
0.1	17.10	.00	128	31.64	23.09	17.10	12.31	9.06
0.2	17.84	0.01	256	33.03	24.10	17.85	12.86	9.46
0.3	18.65	0.04	384	34.59	25.24	18.70	13.46	9.91
0.4	19.54	0.11	512	36.37	26.54	19.66	14.15	10.42
0.5	20.52	0.27	640	38.46	28.06	20.79	14.97	11.02
0.6	21.60	0.58	768	41.03	29.94	22.18	15.97	11.75
0.7	22.80	1.21	896	40.33	30.73	24.01	18.49	14.40
0.8	24.14	2.61	1024	40.13	32.64	26.75	21.94	17.93
0.9	25.65	6.41	1152	46.49	38.48	32.06	27.58	24.05
1	27.36	19.34	1280	65.38	55.11	46.70	42.03	38.30
1.1	29.31	53.19	1408	115.51	97.36	82.51	74.26	67.66

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

LOS	MAXIMUM STOPPED DELAY	QP=1 MAX SV	QP=1 V/C	QP=2 MAX SV	QP=2 V/C	QP=3 MAX SV	QP=3 V/C	QP=4 MAX SV	QP=4 V/C	QP=5 MAX SV	QP=5 V/C
A	5										
	10									407	0.32
B	15							644	0.50	918	0.72
	20					551	0.43	952	0.74	1067	0.83
C	25			357	0.28	942	0.74	1094	0.85	1161	0.91
	30			778	0.61	1102	0.86	1173	0.92	1205	0.94
	35	414	0.32	1076	0.84	1178	0.92	1218	0.95	1250	0.98
D	40	717	0.56	1164	0.91	1221	0.95	1262	0.99		
	45	1122	0.88	1202	0.94	1265	0.99				
	50	1176	0.92	1241	0.97						
	55	1210	0.95	1279	1.00						
E	60	1244	0.97								

AVERAGE STOPPED DELAY FOR:

TABLE A-10

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 120
G/C= 0.5
CAPACITY 1600

V/C	U.D.	R.D.	FLOW RATE	QUALITY OF PROGRESSION 1	2	3	4	5
0	11.40	0.00	0	21.09	15.39	11.40	8.21	6.04
0.1	12.00	.00	160	22.20	16.20	12.00	8.64	6.36
0.2	12.67	0.01	320	23.45	17.11	12.68	9.13	6.72
0.3	13.41	0.03	480	24.87	18.15	13.45	9.68	7.13
0.4	14.25	0.09	640	26.53	19.36	14.34	10.33	7.60
0.5	15.20	0.22	800	28.52	20.81	15.42	11.10	8.17
0.6	16.29	0.46	960	30.98	22.61	16.75	12.06	8.88
0.7	17.54	0.97	1120	31.09	23.69	18.51	14.25	11.11
0.8	19.00	2.11	1280	31.67	25.76	21.11	17.31	14.15
0.9	20.73	5.30	1440	37.74	31.24	26.03	22.39	19.52
1	22.80	17.30	1600	56.14	47.32	40.10	36.09	32.88
1.1	25.33	51.27	1760	107.24	90.39	76.60	68.94	62.81

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

LOS	MAXIMUM STOPPED DELAY	QP=1 MAX SV	QP=1 V/C	QP=2 *MAX SV	QP=2 V/C	QP=3 *MAX SV	QP=3 V/C	QP=4 *MAX SV	QP=4 V/C	QP=5 *MAX SV	QP=5 V/C
A	5				ERR	ERR					
	10				ERR	ERR		559	0.35	1041	0.65
B	15				ERR	738	0.46	1159	0.72	1305	0.82
	20			710	0.44	1212	0.76	1365	0.85	1446	0.90
C	25	492	0.31	1221	0.76	1406	0.88	1471	0.92	1506	0.94
	30	896	0.56	1404	0.88	1485	0.93	1529	0.96	1565	0.98
	35	1368	0.85	1477	0.92	1542	0.96	1587	0.99		
D	40	1460	0.91	1527	0.95	1599	1.00		ERR		
	45	1503	0.94	1577	0.99				ERR		
	50	1547	0.97						ERR		
	55	1590	0.99						ERR		
E	60	ERR	ERR						ERR		

AVERAGE STOPPED DELAY FOR:

TABLE A-11

SIGNAL TYPE= PRETIMED
SAT FLOW= 3200
CYCLE= 120
G/C= 0.6
CAPACITY 1920

V/C	U.D.	R.D.	FLOW RATE	QUALITY OF PROGRESSION 1	2	3	4	5
0	7.30	0.00	0	13.50	9.85	7.30	5.25	3.87
0.1	7.76	.00	192	14.36	10.48	7.76	5.59	4.11
0.2	8.29	0.01	384	15.35	11.20	8.30	5.97	4.40
0.3	8.90	0.03	576	16.51	12.05	8.93	6.43	4.73
0.4	9.60	0.08	768	17.90	13.06	9.68	6.97	5.13
0.5	10.42	0.18	960	19.61	14.31	10.60	7.63	5.62
0.6	11.40	0.39	1152	21.80	15.91	11.79	8.49	6.25
0.7	12.58	0.81	1344	22.50	17.14	13.39	10.31	8.03
0.8	14.03	1.77	1536	23.71	19.28	15.81	12.96	10.59
0.9	15.86	4.52	1728	29.56	24.46	20.39	17.53	15.29
1	18.24	15.79	1920	47.65	40.16	34.03	30.63	27.91
1.1	21.46	49.91	2112	99.92	84.22	71.37	64.24	58.53

SERVICE VOLUME AND VOLUME/CAPACITY RATIO

LOS	MAXIMUM STOPPED DELAY	QP=1 MAX SV	QP=1 V/C	QP=2 *MAX SV	QP=2 V/C	QP=3 *MAX SV	QP=3 V/C	QP=4 *MAX SV	QP=4 V/C	QP=5 *MAX SV	QP=5 V/C
A	5									706	0.37
	10			46	0.02	835	0.43	1311	0.68	1492	0.78
B	15	316	0.16	1043	0.54	1472	0.77	1622	0.84	1716	0.89
	20	994	0.52	1563	0.81	1712	0.89	1764	0.92	1800	0.94
C	25	1578	0.82	1735	0.90	1793	0.93	1837	0.96	1876	0.98
	30	1733	0.90	1796	0.94	1863	0.97	1911	1.00		
	35	1786	0.93	1857	0.97						
D	40	1839	0.96	1918	1.00						
	45	1892	0.99								
	50										
	55										
E	60										

Saturation Flows of Exclusive Double Left-Turn Lanes

ROBERT W. STOKES, CARROLL J. MESSER, AND VERGIL G. STOVER

The objectives of this study were to develop estimates of the saturation flows of exclusive double left-turn lanes, and to investigate the physical and operating characteristics of the intersection that may affect left-turn saturation flows. The results are based on observations of 3,458 completed left turns from exclusive double left-turn lanes on 14 intersection approaches in Austin, College Station, and Houston, Texas. Based on the results of this study and a review of the data from a limited number of related studies, an average double left-turn saturation flow rate of approximately 1,600 vehicles per hour of green per lane would appear to be a reasonable value for most planning applications. This flow rate can be assumed to be achieved for the third vehicle in the queue and beyond. Also, this flow rate appears to be applicable for mixed traffic conditions in which heavy vehicles constitute as much as 3 to 5 percent of the left-turn traffic volume.

Texas Transportation Institute, Texas A&M University System, College Station, Tex. 77843.

The continuing emphasis on obtaining a more efficient utilization of the existing urban street infrastructure suggests that applications of the double left-turn lane concept will become increasingly widespread. Reliable estimates of double left-turn saturation flows have several important applications in traffic

management and design. Such applications include the following *(1)*:

- Determination of optimum signal timing and phasing arrangements,
- Estimation of average queue lengths for use in the design of left-turn storage bays,
- Estimation of double left-turn lane capacity, and
- Estimation of the average and maximum delays experienced by left-turning vehicles.

Although double left-turn lanes have been in use for a number of years, basic research on the saturation flows of such facilities has been extremely limited. Consequently, current methods for estimating the capacities of double left-turn lanes are based on minimal amounts of data and none has been widely accepted as being truly representative of real-world conditions.

The capacity estimates derived from the limited data available on double left-turn lanes suggest a substantial range in flow rates. This range could be attributed to various physical, operating, and/or environmental factors that may vary within or between intersection approaches. This paper summarizes the results of a study directed at (*a*) estimating exclusive double left-turn lane capacity, which is easily computed from saturation flow; and (*b*) quantifying the relationships between double left-turn lane saturation flow and the physical and operating characteristics of the intersections studied *(2)*.

The use of the phrase "exclusive double left-turn lanes" in this paper refers to two contiguous lanes of an intersection approach that are designed solely for left-turning vehicles and protected from opposing through- and cross-traffic with separate signal phasing. In the study, double left-turn movement capacity is dealt with on a microscopic level; that is, in terms of the time headways of individual left-turning vehicles.

OBJECTIVES

The primary objective of the study was to develop reliable estimates of exclusive double left-turn saturation flows. A secondary objective was to attempt to relate these flows to the physical and operating characteristics of the intersection approaches studied. Specific objectives were as follows:

1. Record, with an accuracy of ±0.1 sec, the time spacing between consecutive left-turning vehicles entering a signalized intersection from an exclusive double left-turn lane during saturated conditions (i.e., for queue lengths of five or more vehicles per lane).
2. Examine the variability in these time spacings between and within intersection approaches.
3. If possible, relate these time spacings and their variability in terms of the following physical and operating characteristics of the intersection approaches studied: (*a*) Width of turn lanes (ft); (*b*) Turn radius (ft); (*c*) Approach grade (percent); (*d*) Turn-bay storage length (ft); (*e*) Turn-bay taper length (ft); (*f*) Turn-bay queue lengths (veh/lane); (*g*) Turn-lane blockages, (*h*) Turn-lane green time (sec); and (*i*) Percentage of heavy vehicles in the left-turn traffic stream (i.e., percentage of vehicles with more than four tires).

PREVIOUS RESEARCH

Double left-turn lanes have been in use for a number of years. However, basic research on the saturation (capacity) flows of such facilities has been extremely limited. A literature search revealed only seven references *(3–9)* on the subject of double left-turn lane capacity. Moreover, only five of the references treated double left-turn lanes in detail. With the exception of Kunzman's limited work in 1978 *(9)*, the literature search indicated that no fundamental research on double left-turn movement capacity has appeared in the literature in nearly 15 years.

Table 1 gives a summary of the estimates of double left-turn saturation flows reported in the research identified through the literature search. Also given is a summary of the efficiencies of double left-turn movements relative to single left-turns and straight-through movements. The average double left-turn saturation flows for the limited data available and given in Table 1 span a considerable range of values. However, with the exception of Ray's data *(4)*, and the Highway Capacity Manual (HCM) data for through-movements *(5)*, the efficiency factors appear to be fairly uniform. Also, note that with the exception of the HCM data, the left-turn flow rates do not appear to be substantially lower than through-lane flow rates. This suggests that the flows for all movements on a particular approach are equally affected by the physical and operating characteristics of the intersection. That is, those approaches with relatively low double left-turn flows also appear to have proportionately lower single left-turn and straight-through flows.

DATA COLLECTION AND ANALYSIS

Based on an examination of the possible methods for obtaining the data required, it was decided that the most satisfactory and economical procedure, from both field study and data analysis time standpoints, was a combination of photographic and field measurement techniques.

By using time-lapse photography, more than 3,400 completed left turns were recorded on 14 intersection approaches with exclusive double left-turn lanes in College Station, Austin, and Houston, Texas. These three cities were selected for study because they represented a cross-section of Texas cities in terms of population and traffic conditions.

A filming speed of 9 frames/sec was used because this speed made it possible to estimate left-turn departure headways to within 0.1 sec. A total of approximately 4 hours of peak-period data was collected. This was considered a sufficient sample for estimating double left-turn saturation flows and for making statistical inferences concerning the factors affecting saturation flow.

The Statistical Analysis System (SAS) Computer Program Package was used in the statistical analysis phases of the study *(10)*. This package has been extensively tested and has been widely accepted for statistical analyses.

The sample data were used to test specific statistical hypotheses concerning the following:

1. The saturation flow region of the queue,
2. Variability in the estimates of left-turn departure headways, and

TABLE 1 DOUBLE LEFT-TURN SATURATION FLOWS AND RELATIVE EFFICIENCY FACTORS

Source and Conditions	Double Left-Turn Saturation Flow (vphg)			Double Left-Turn Movement Efficiency Factor[a]	
	Inside Lane	Outside Lane	Average	Relative to Single Left-Turn	Relative to Through Movement
Capelle and Pinnell (3)					
Permissive Double Left-Turn[b]	1500	1636	1568	0.91	0.91
Ray (4)					
Permissive Double Left-Turn	1240	1230	1235	0.75	-[c]
Exclusive Double Left-Turn	1375	1315	1345	0.82	-
HCM (5)					
10-foot Lanes	1200	960	1080	0.90	0.72[d]
11-foot Lanes	1320	1056	1188	0.90	0.79[d]
12-foot Lanes	1440	1152	1296	0.90	0.86[d]
Assmus (7)					
Exclusive Double Left-Turn	1540	1550	1545	-	0.97
Kunzman (9)					
Queue $\leq$ 4 veh/lane	-	-	1439	0.96	0.90
Queue $\geq$ 5 veh/lane	-	-	1581	0.92	0.93
All Queue Lengths	-	-	1523	0.93	0.91

[a] Efficiency factor defined as the ratio of average double left-turn flow to average single left-turn flow or average straight-through flow as indicated; all flows are per-lane flows.
[b] Left-Turn option in outside lane.
[c] Data not available.
[d] Assumes straight-through flow of 1500 vphg.

3. Some of the factors that may affect left-turn departure headways.

The statistical hypotheses were tested at the 5 percent significance level. The statistical procedures employed are discussed briefly in the following subsections.

Saturation Flow Region of the Queue

The initial task in estimating double left-turn saturation flows was to identify the portion of the queue, in terms of vehicle storage positions, for which departure headways could be assumed to be uniform. Two basic statistical procedures were employed to accomplish this task: analysis of means and regression analysis.

In the analysis of means procedure, average left-turn headways were calculated by turn lane (Figure 1) for the entire sample (i.e., for all intersection approaches), by turn lane and approach, and by turn lane and city. The average headways were calculated for all vehicles observed entering the intersections, for selected segments of the queue, and by individual storage positions. In calculating the means by queue segments, the following regions of the queue were considered:

$$1 \leq n_i \leq n_{k(i)}, \ldots, 5 \leq n_i \leq n_{k(i)}$$

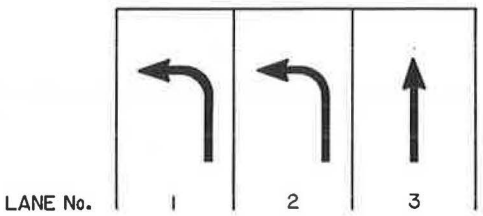


FIGURE 1 Lane numbering scheme.

where n_i is the queue storage position for left-turning vehicles in Turn Lane i, for i = 1,2 (Figure 1); and $n_{k(i)}$ is the total number of vehicles queued in Lane i at the start of the left-turn green phase. Duncan's Multiple Range Test was used to test the hypothesis that average left-turn headways do not differ by vehicle storage position. Table 2 presents an aggregate summary of average left-turn headways for the total sample.

The regression analyses involved evaluating the following models:

$$T_{c(i)} = B_0 + B_1 n_I + e_i \quad (1)$$

$$h_{I,(i)} = B^*_0 + B^*_1 (1/n_i) + e^*_i \quad (2)$$

TABLE 2 AVERAGE LEFT-TURN DEPARTURE HEADWAYS FOR THE TOTAL SAMPLE

Data Set and Lane	Sample Size	Mean Headway (sec/veh)	Standard Deviation	Standard Error of Mean
All Vehicles				
Lane 1	1707	2.4	1.34	0.03
Lane 2	1751	2.4	1.29	0.03
$1 \leq n_i \leq n_{k(i)}$				
Lane 1	1500	2.3	0.81	0.02
Lane 2	1598	2.3	1.04	0.03
$2 \leq n_i \leq n_{k(i)}$				
Lane 1	1249	2.2	0.64	0.02
Lane 2	1347	2.2	0.98	0.03
$3 \leq n_i \leq n_{k(i)}$				
Lane 1	1000	2.1	0.60	0.02
Lane 2	1101	2.1	1.00	0.03
$4 \leq n_i \leq n_{k(i)}$				
Lane 1	752	2.0	0.58	0.02
Lane 2	859	2.1	1.09	0.04
$5 \leq n_i \leq n_{k(i)}$				
Lane 1	509	2.0	0.60	0.03
Lane 2	623	2.1	1.17	0.05

where

$T_{c(i)}$ = time (sec) after the start of the left-turn green phase for the rear axle of the left-turning vehicle in queue storage position n_i to cross the intersection stop line from Turn Lane i (i = 1,2);
B_0, B_1, B^*_0, B^*_1 = regression coefficients;
e_i, e^*_i = random error terms; and
$h_{L(i)}$ = left-turn headway (sec) for vehicles in Turn Lane i (i = 1,2).

In Equation 1, the parameter of interest is B_1, where the sample estimate of B_1, denoted by b_1, is an estimate of departure headway (sec/veh). The parameters of the model given in Equation 1 were estimated from the entire sample, as well as from subsets of the sample. The subsets of the sample used to estimate the parameters of the regression models were identical to those used in the analysis of means.

A less direct analysis of the saturation flow region of the queue was performed by evaluating regression models of the form given in Equation 2. The intent of the analyses involving Equation 2 was to find n_i for which the hypothesis $B_1 = 0$ could not be rejected. In addition, the corresponding b_0's should then be estimates of minimum average departure headways.

Variability in Estimates of Left-Turn Departure Headways

The purpose of the second phase of the analysis was to determine whether the average left-turn headways for the saturation flow region of the queue differed significantly between or within intersection approaches, and/or by turn lane. The following three procedures were used to investigate the variability exhibited by the estimates of left-turn departure headways.

1. Duncan's Multiple Range Test was used to test the hypothesis that average left-turn departure headways do not differ significantly between intersection approaches.
2. A series of two-sample *t*-tests was performed to test the differences in departure headways between cities. The tests were performed on a per-lane basis by using the *t*-test (TTEST) procedure of the SAS Computer Program Package *(10)*. The TTEST procedure tests the significance of the difference between means from two independent samples. The TTEST procedure also performs an *F*-test on the hypothesis that the variances of the two samples are equal. The procedure then calculates two *t*-statistics: the usual two-sample *t* that assumes equal variances, and an approximate *t* that does not assume equal variances *(11)*.
3. A paired *t*-test was used to test the hypothesis that average left-turn departure headways do not differ between the two lanes of the double left-turn movements within each of the three cities studied.

Factors Affecting Left-Turn Departure Headways

The intent of the final phase of the analyses was to examine some of the possible sources of the variability observed in the left-turn departure headway estimates. The potential headway-influencing factors investigated are summarized in Table 3.

The initial step in investigating the effects of the numeric factors (i.e., factors other than lane blockages) in the sample was to calculate the simple correlation coefficients (*r*) between the departure headways and the individual factors. The correlation (CORR) procedure of the SAS Computer Program Package was used to compute the sample correlation coefficients *(10)*.

Note in Table 3 that in the final analyses only three levels of lane blockages were considered. Specifically, only the blockage, no blockage, and blockage code 23 conditions were con-

TABLE 3 HEADWAY-INFLUENCING FACTORS INVESTIGATED

Factor	Contraction	Definition
Lane blockage[a]	NOBLK	No lane blockages (code 00)
	BLOCK	Lane blockage codes 12, 21, and 23 combined
	BLK23	Blockage code 23
Headway compression[b]	COMPRES1	Ratio of total queue length (veh) to available green time (sec)
	COMPRES2	for Lane 1 and Lane 2, respectively
Approach grade	GRADE	Approach grade (%)
Green time	GRNTIME	Left-turn green time (sec)
Percent heavy vehicles	PERHV1	Percentage heavy vehicles (veh with more than four
	PERHV2	tires) in Lane 1, Lane 2, and Lanes 1 and 2
	PERHVTOT	combined, respectively
Turn radius	RADIUS	Simple turn radius (ft)
Turn bay storage length	STORAGE	Turn-bay storage length (ft)
Turn bay taper length	TAPER	Turn-bay taper length (ft)
Turn lane width	WIDTH1	Width (ft) of Turn Lane 1
	WIDTH2	Width of Lane 2
	TOTWIDTH	Width of Lanes 1 and 2 combined

[a]Code 00 = no blockage observed, code 12 = Lane 1 blocked by Lane 2, and so forth. See Figure 1 for lane numbering scheme.
[b]As the demand (queue length) per cycle increases relative to capacity (left-turn green time), there may be some compression, or shortening, of left-turn departure headways as vehicles attempt to fully utilize available capacity. It would be expected that the effects of this compression factor be more pronounced in those situations in which the available green time is not sufficient to fully serve the queue of vehicles waiting on the approach.

sidered in the analyses. These revised definitions of the blockage codes were formulated to facilitate the testing of statistical hypotheses suggested by the data. The effects of turn lane blockages on left-turn departure headways were evaluated by testing two hypotheses:

- The average departure headways for blockage codes 12, 21, and 23 combined do not differ significantly from those for the no-blockage condition (code 00).
- The average departure headways for blockage code 23 do not differ significantly from those for the no-blockage condition (code 00).

The first hypothesis was tested for each lane of the College Station and Houston sites. The second hypothesis was tested for Lane 2 of the College Station and Houston sites individually, as well as for Lane 2 of the two cities combined. No turn lane blockages were observed at the Austin sites. The *t*-test (TTEST) procedure of the SAS Computer Program Package was used to perform the statistical tests *(10)*.

RESULTS

Saturation Flow Region of the Queue

Three slightly different estimates of the saturation flow region of the queue were developed. The estimates of the saturation flow region of the queue and the corresponding headway estimates—developed from the analysis of means and the regression models given in Equation 1—were in fairly close agreement. In terms of the precision of the departure headway estimates, the estimates obtained from the regression models given in Equation 2 were inferior to those obtained from the other two estimation procedures. The estimates obtained from the analysis of means were selected as the best estimates. The results of the analysis of means suggest that the saturation flow region of the queue can be defined by the region $3 \leq n_i \leq n_{k(i)}$. The decision to use the estimates obtained from the analysis of means was based on the following considerations.

First, from a theoretical standpoint, it can be argued that the regression estimates (Equation 1) of departure headways should be more precise than those obtained from the analysis of means. However, it appears that at least one of the basic assumptions of regression analysis was not satisfied. Specifically, it was found that the variance of the error term in the regression models was not constant across the sample. The procedure used to test the constancy of the error variance was to

1. Array the observations by queue storage position and lane,
2. Divide the total observations into two equal data sets for each lane,
3. Fit separate regression functions to each half of the total observations,
4. Calculate the mean square errors (MSEs) for each, and
5. Test for equality of the error variances by the *F*-test.

The resulting variance ratios were significant at the 5 percent level.

Thus, evaluation of the constancy of the error variance suggests that the regression estimates of the departure headways may be biased. The data in Table 4 demonstrate the nature of the suspected bias. Note that, with the exception of Lane 2 for the Austin sites, the estimated headways (i.e., the slopes) are all less than or equal to 2.0 sec. The regression estimates suggest average flow rates in excess of 1,800 vehicles per hour of green per lane (vphgl) for the majority of the study sites. Relative to generally accepted straight-through flow rates of 1,700 to 1,800 vphgl, the regression models appear to have underestimated average left-turn departure headways. (See paper by Stokes, Stover, and Messer elsewhere in this Record for discussion.)

Variability in Departure Headways

The results of Duncan's Test indicated significant differences in the average left-turn headways between the intersection approaches studied. However, there was a general pattern to the Duncan rankings of the headways; that is, the headways were generally grouped by city size. Specifically, the Houston sites (the large city studied) tended to exhibit significantly shorter average headways than did the College Station and Austin sites (the small- and medium-sized cities studied).

Table 5 presents the 95 percent confidence intervals for the average left-turn departure headways for the saturation flow region of the queue [i.e., for $3 \leq n_i \leq n_{k(i)}$] for each of the three cities studied. *F*-tests were performed on the hypothesis that the average headways do not differ between cities. The *F*-tests

TABLE 4 EQUATION 1 REGRESSION MODELS BY CITY FOR $3 \leq n_i \leq n_{k(i)}$

City and Lane	Sample Size	Intercept Estimate (sec)	Intercept Std. Err.	Slope Estimate (sec/veh)	Slope Std. Err.	Mean Sq. Err. (MSE)	r^2
Austin							
Lane 1	175	2.2	0.37	1.9	0.06	2.79	0.84
Lane 2	180	1.3	0.31	2.2	0.05	2.09	0.91
College Station							
Lane 1	486	2.5	0.20	1.9	0.04	2.57	0.86
Lane 2	592	2.2	0.26	2.0	0.04	5.07	0.80
Houston							
Lane 1	339	2.4	0.22	1.8	0.05	1.04	0.79
Lane 2	329	2.5	0.24	1.7	0.06	1.33	0.74

TABLE 5 95 PERCENT CONFIDENCE INTERVALS FOR AVERAGE LEFT-TURN HEADWAYS (h_L) BY CITY FOR $3 \leq n_i \leq n_{k(i)}$

	95% Confidence Intervals for Average Headways (sec/veh)		
Lane	Austin	College Station	Houston
1	$2.1 \leq \bar{h}_L \leq 2.3$	$2.0 \leq \bar{h}_L \leq 2.2$	$1.9 \leq \bar{h}_L \leq 2.1$
2	$2.1 \leq \bar{h}_L \leq 2.3$	$2.1 \leq \bar{h}_L \leq 2.3$	$1.8 \leq \bar{h}_L \leq 2.0$

indicated significant differences between the average headways of the double left-turn lanes of the three cities studied. A series of *t*-tests was performed to test the differences in average left-turn headways between cities. The following conclusions can be drawn from the tests:

- The average departure headways for the Austin and College Station sites are not significantly different for either turn lane for the saturation flow region of the queue. (Lane 1: t = 1.92 and p-value = 0.06. Lane 2: t = 0.81 and p-value = 0.42).
- The average departure headways for the Houston sites are significantly shorter than those for the Austin sites for each turn lane for the saturation flow region of the queue. (Lane 1: t = 3.67 and p-value < 0.01. Lane 2: t = 4.78 and p-value < 0.01).
- The average departure headways for the Houston sites are significantly shorter than those for the College Station sites for each turn lane for the saturation flow region of the queue. (Lane 1: t = 2.22 and p-value = 0.03. Lane 2: t = 3.72 and p-value < 0.01).

The differences in the average headways between the two lanes of the double left-turn movements within each city were also examined. As indicated by the data in Table 6, the average differences *(D)* between the Lane 1 and Lane 2 headways were generally positive, indicating that the Lane 1 headways tended to be slightly longer than the Lane 2 headways. Note, however, that none of the differences were significant at the 5 percent level. Thus, it can be concluded that any differences in the departure headways for vehicles in Lane 1 and Lane 2 within each city are not large enough to be detected with the given sample sizes.

Factors Affecting Departure Headways

Two basic statistical procedures were used to investigate the factors in the sample that might have an effect on left-turn saturation flows. Correlation analyses were used to assess the effects of the numeric factors in the sample. The effects of turn lane blockages were evaluated by using analysis-of-means procedures.

TABLE 6 *t*-TEST OF DIFFERENCES IN LEFT-TURN HEADWAYS BETWEEN LANES BY CITY FOR $3 \leq n_i \leq n_{k(i)}$

City	Variable	Mean	Std. Dev.	Std. Err. of Mean	t	Prob > |t|	Significant at 5% Level
Austin	D[a]	-0.08	1.30	0.10	-0.81	0.4202	NO
College Station	D	0.10	1.79	0.08	1.29	0.1967	NO
Houston	D	0.07	0.95	0.05	1.19	0.2341	NO

[a] $D = h_{L(1)} - h_{L(2)}$.

Table 7 gives a summary of those factors that exhibited significant linear correlations with departure headways in the saturation flow region of the queue. Note in the table that approach grade (GRADE), the headway compression factors (COMPRES1 and COMPRES2), and average left-turn green time (GRNTIME) exhibit significant correlations with both the Lane 1 and Lane 2 headways. Figure 2 shows these general relationships in terms of saturation flows.

TABLE 7 SIGNIFICANT[a] SIMPLE CORRELATION COEFFICIENTS (r) FOR THE NUMERIC FACTORS IN THE SAMPLE FOR $3 \leq n_i \leq n_{k(i)}$

	Lane 1 Headway		Lane 2 Headway	
Factor[b]	r	Prob > |r|[c]	r	Prob > |r|
TAPER	0.0885	(0.0082)	-[d]	
STORAGE	0.0833	(0.0084)	-	
GRADE	0.0952	(0.0026)	0.0965	(0.0013)
PERHV1	0.0778	(0.0138)	-	
COMPRES1	-0.1126	(0.0004)	-0.1758	(0.0001)
COMPRES2	-0.0897	(0.0048)	-0.1645	(0.0001)
GRNTIME	0.0977	(0.0021)	0.2235	(0.0001)
WIDTH1	-		-0.0915	(0.0024)
WIDTH2	-		-0.0827	(0.0061)
TOTWIDTH	-		-0.0941	(0.0018)

[a] Significance level = 0.05.
[b] See Table 3 for definitions.
[c] For hypothesis r = 0.
[d] Not significant at 5% level.

The positive relationship between saturation flow and approach grade (Figure 2*a*) is not surprising. It appears reasonable to expect departure headways to increase and saturation flows to decrease as the approach grade increases. However, given the limited range of approach grades in the sample (Figure 2*a*), the exact nature of the relationship remains open to question.

The relationships between saturation flow and the headway compression and green time factors (Figures 2*b*–2*d*) are not as easily interpreted as the other relationships examined. Although it appears reasonable to expect flow rates to increase as the turn lanes become busier, one should not lose sight of the implications of this assumption. Specifically, the effects of the compression and green time factors would appear to be related to driving behavior, not to quantifiable features of the turn lanes. That is, the observed effects of the compression and green time factors would appear to imply that drivers recognize

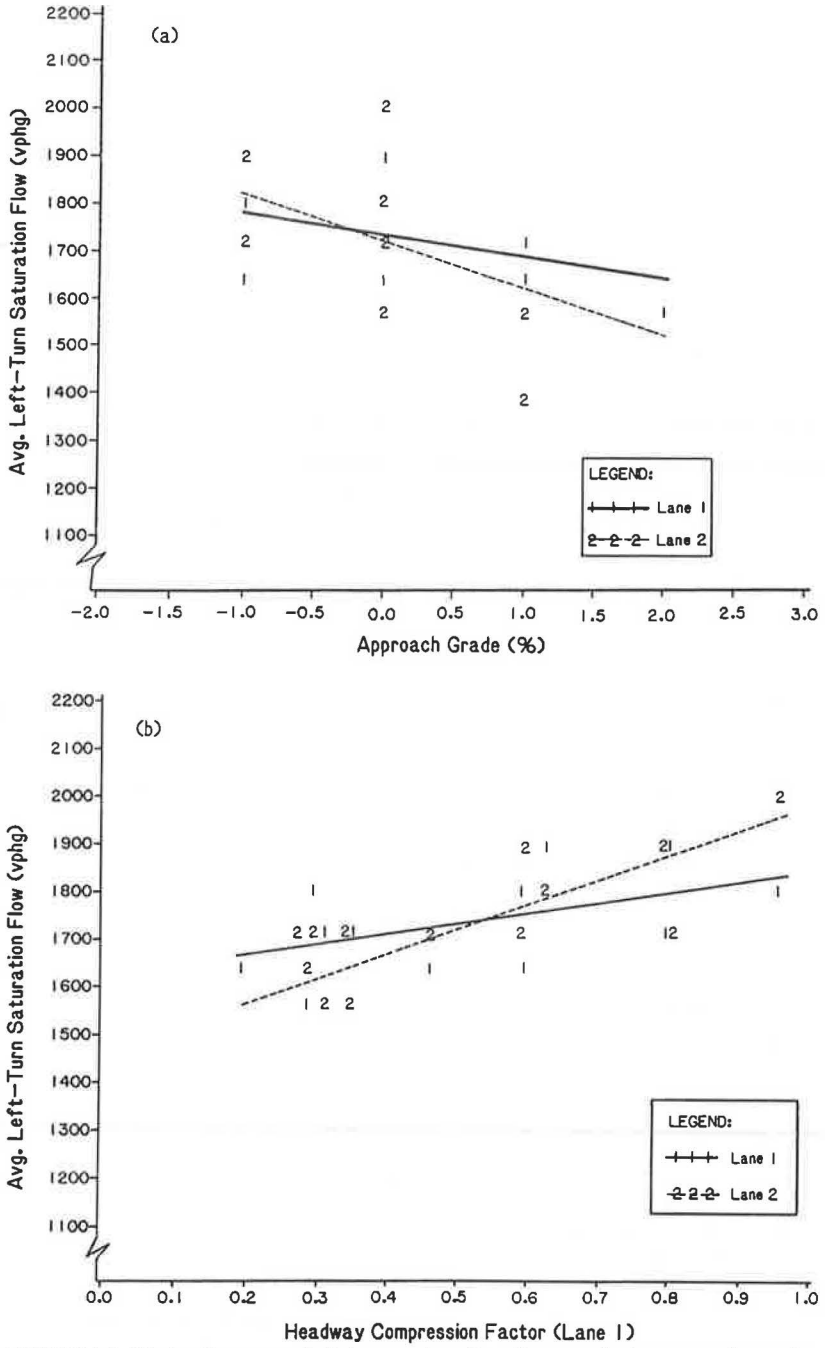
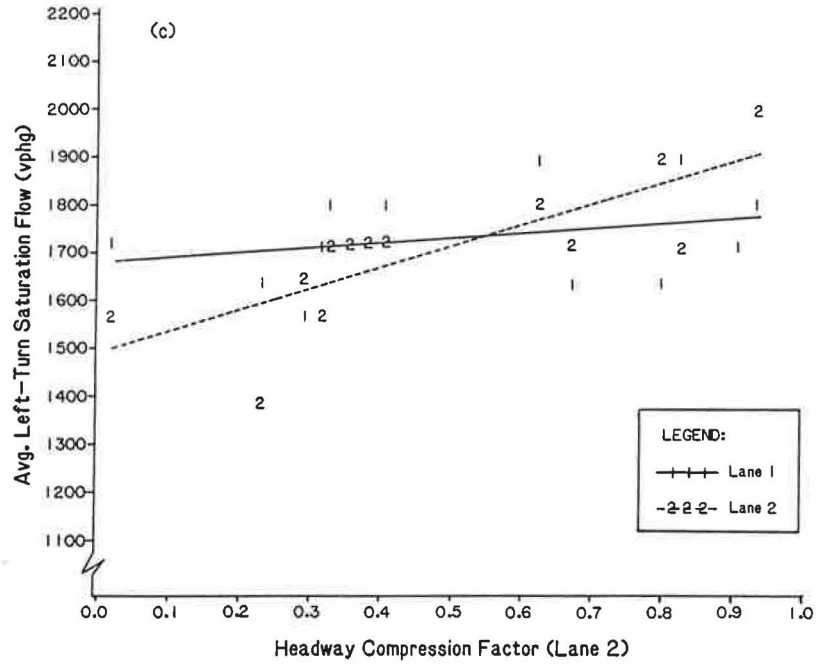


FIGURE 2 Plots of average left-turn saturation flow against approach grade, headway compression factors, and average left-turn green time.

that available green time is short (relative to the demand on the approach) and that they respond by assuming a relatively more aggressive driving posture. If the compression and green time factors are related to driving behavior, the magnitudes of the effects of these factors are probably highly variable.

The *t*-test was used to test the hypothesis that turn lane blockages have no effect on the average left-turn departure headways for the saturation flow region of the queue. The results of the tests for the saturation flow region of the queue are summarized as follows.

- Turn lane blockages have no significant effect on the left-turn departure headways of Lane 1 of the College Station sites ($t = -1.01$ and p-value = 0.31).
- Turn lane blockages have a significant effect on the left-turn departure headways of Lane 1 of the Houston sites ($t = 2.25$ and p-value = 0.02). Specifically, the departure headways of Lane 1 of the Houston sites for lane blockage conditions appear to be shorter than in the cases in which no lane blockage is encountered.
- Turn lane blockages have a significant effect on the left-



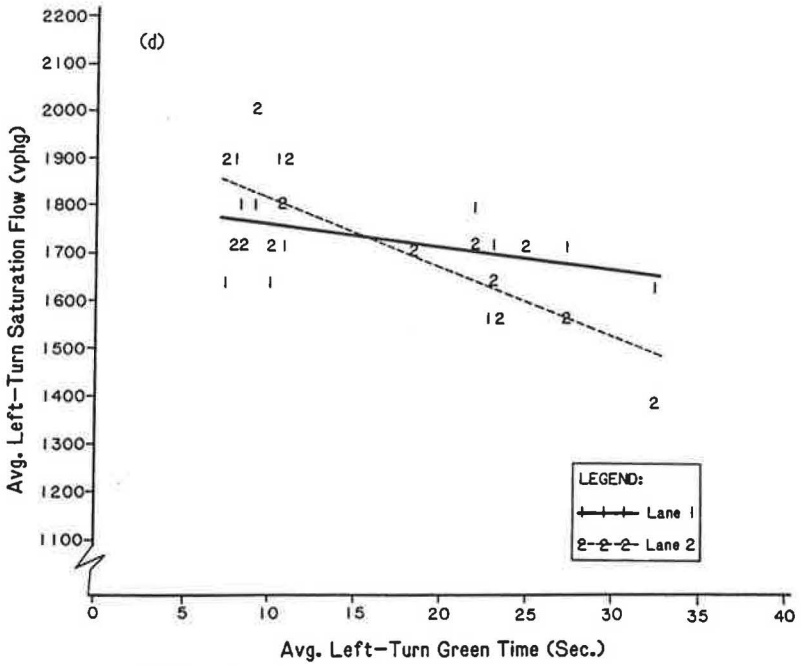


FIGURE 2 *continued.*

turn departure headways of Lane 2 of the College Station sites ($t = -2.20$ and p-value = 0.03). The departure headways for the no-blockage conditions appear to be shorter than in the cases in which lane blockages are encountered.

- Turn lane blockages have no significant effect on the left-turn departure headways of Lane 2 of the Houston sites ($t = -0.33$ and p-value = 0.75).
- In the case in which Lane 2 of the double left-turn movement is blocked by the adjacent through lane (Lane 3), the left-turn departure headways for Lane 2 are not significantly different from those in which no blockage occurs, for either the Houston or College Station sites (Houston: $t = -0.74$ and p-value = 0.47; College Station: $t = -1.59$ and p-value = 0.12). For the two cities combined, the departure headways for Lane 2 for the no-blockage condition do not appear to be significantly different from the departure headways for the case in which Lane 2 is blocked by the adjacent through lane ($t = -1.88$ and p-value = 0.06).

SUMMARY

It is indicated in the analyses that saturation flow for each lane of a double left-turn movement is obtained after the second vehicle in the queue. Vehicles, after the second in line, enter the intersection at the saturation flow rate until at least the last vehicle in the initial queue has entered the intersection or until the end of the effective left-turn green time.

It is suggested by the analyses that average left-turn departure headways vary significantly between the intersection approaches studied. However, much of the variability in the departure headways can be accounted for if the headways are grouped according to city size. Also, the study results indicate that, within each city, the departure headways do not differ significantly between the two lanes of a double left-turn movement.

Table 8 gives a summary of the 95 percent confidence intervals for average left-turn saturation flows ($\bar{s}_L$) estimated for the three Texas cities studied. As indicated by the data in the table, the left-turn saturation flow estimates developed in this study are generally higher than those that have been reported by other researchers (see Table 1). The average flow rates given for the Houston sites are on the same order of magnitude as the flows commonly reported for straight-through movements.

TABLE 8 95 PERCENT CONFIDENCE INTERVALS FOR AVERAGE LEFT-TURN SATURATION FLOWS (vphgl) BY CITY FOR $3 \leq n_i \leq n_{k(i)}$

	95% Confidence Intervals[a] for Average Left-Turn Saturation Flow ($\bar{s}_L$)		
Lane	Austin	College Station	Houston
1	$1565 \leq \bar{s}_L \leq 1714$	$1636 \leq \bar{s}_L \leq 1800$	$1714 \leq \bar{s}_L \leq 1895$
2	$1565 \leq \bar{s}_L \leq 1714$	$1565 \leq \bar{s}_L \leq 1714$	$1800 \leq \bar{s}_L \leq 2000$
Grand Mean	1636	1636	1800

[a] Calculated from $(1/\bar{h}_L) \times 3600$, using $\bar{h}_L$'s from Table 5.

The factors affecting left-turn departure headways were the subjects of some exploratory analyses. Although the results of these preliminary investigations were inconclusive, they suggest several complex relationships in the operating characteristics of double left-turn lanes. It is indicated by the investigations that approach grade, the headway compression factors, and average left-turn green time exhibit significant linear correlations with both the Lane 1 and Lane 2 departure headways.

Two general, and apparently contradictory, trends regarding the effects of turn lane blockages on left-turn departure headways are suggested by the study results. First, it appears that the effects of lane blockages vary by lane and city size. For College Station (the small city in the study), the effects of the lane blockages appear to be focused on the departure headways of Lane 2 (outside lane), where the blockages tend to increase the departure headways. For the large city in the sample (i.e., Houston), the effects of the lane blockages appear to be focused on the Lane 1 (inside lane) departure headways, where the blockages tend to reduce the departure headways. Second, for the specific lane blockage condition in which Lane 2 of the double left-turn movement is blocked by the adjacent through lane, it appears that this blockage condition has no significant effect on the Lane 2 headways for either the College Station or Houston sites individually, or the College Station and Houston sites combined.

Based on the results of this study and a review of the data from a limited number of related studies, an average double left-turn saturation flow rate of approximately 1,600 vphgl would appear to be a reasonable value for most planning applications. This flow rate can be assumed to be achieved for the third vehicle in the queue and beyond. Also, this flow rate appears to be applicable for mixed traffic conditions in which heavy vehicles constitute as much as 3 to 5 percent of the left-turn traffic volumes.

The double left-turn saturation flows observed at the Houston sites were approximately 1,800 vphgl. When compared with the flow rates typically reported in the literature for left-turn and straight-through movements, the Houston flow rates appear to be high. However, the Houston flow rates may be indicative of the maximum flows that can be realized on double left-turn lanes in an urban environment.

It is suggested by the results of this study that average double left-turn saturation flows may be substantially higher than previously believed. However, the results are based on data from only three cities in Texas. Consequently, in assessing a site-specific problem, engineers and planners may find it useful to collect local data to spot-check the applicability of the flow rates reported in this study.

ACKNOWLEDGMENTS

The authors wish to thank the following individuals for their contributions to this study: D. L. Pugh, J. deJong, D. A. Maxwell, L. J. Ringer, and A. M. Elmquist, all with Texas A&M University; D. W. Hall, City of Austin; J. R. Black, City of College Station; and W. E. Hensch, City of Houston. The authors remain solely responsible for the contents of this paper.

REFERENCES

1. P. G. Michalopoulos, J. O'Connor, and S. M. Novoa. Estimation of Left-Turn Saturation Flows. In *Transportation Research Record 667*, TRB, National Research Council, Washington, D.C., 1978, pp. 35–41.
2. R. W. Stokes. *Saturation Flows of Exclusive Double Left-Turn Lanes*. Ph.D. dissertation. Texas A&M University, College Station, 1984.
3. D. G. Capelle and C. Pinnell. Capacity Study of Signalized Diamond Interchanges. *Bulletin 291*, HRB, National Research Council, Washington, D.C., 1961, pp. 1–25.
4. J. C. Ray. Two Lane Left-Turns Studied at Signalized Intersections. *Traffic Engineering*, Vol. 35, 1965, pp. 17–19, 58.
5. *Special Report 87: Highway Capacity Manual—1965*. HRB, National Research Council, Washington, D.C., 1965, 411 pp.
6. J. E. Leisch. Capacity Analysis Techniques for Design of Signalized Intersections. *Public Roads*, Vol. 34, 1967, pp. 171–209.
7. W. E. Assmus. *Operational Performance of Exclusive Double*

Left-Turn Lanes. M.S. thesis. Northwestern University, Evanston, Ill., 1970.
8. Institute of Traffic Engineers. The Use and Effectiveness of Double Left-Turn Movements. *Traffic Engineering*, Vol. 45, 1975, pp. 52–57.
9. W. Kunzman. Another Look at Signalized Intersection Capacity. *ITE Journal*, Vol. 48, 1978, pp. 12–15.
10. *SAS User's Guide: Statistics*. SAS Institute Inc., Cary, N.C., 1982.
11. R. J. Freund and R. C. Littell. *SAS for Linear Models: A Guide to the ANOVA and GLM Procedures*. SAS Institute, Inc., Cary, N.C., 1981.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Use and Effectiveness of Simple Linear Regression To Estimate Saturation Flows at Signalized Intersections

ROBERT W. STOKES, VERGIL G. STOVER, AND CARROLL J. MESSER

In this paper, data from 14 intersection approaches with exclusive double left-turn lanes in three Texas cities are used to illustrate the use of simple linear regression to estimate saturation flows. Some theoretical considerations and potential bias in saturation flows estimated by simple linear regression are also briefly discussed. Average left-turn saturation flows in excess of 1,800 vehicles per hour of green per lane for the majority of the study sites within each city are suggested by the regression estimates. Relative to generally accepted straight-through flow rates, the regression models appear to have substantially underestimated average left-turn departure headways.

When a queue of vehicles is released by a traffic signal, the departure flow rate quickly increases until after the first few vehicles, when a uniform average departure rate is reached. This uniform departure rate is called the saturation flow rate of the intersection approach. Because the flow at signalized intersections is controlled by the amount of green time allotted, saturation flow under these conditions is defined as the flow rate that would result if there were a continuous queue of vehicles and they were given 100 percent green time *(1)*. Saturation flow is generally expressed in vehicles per hour of green time (vphg).

A method commonly used to estimate saturation flows at signalized intersections is the headway method. In this method, interarrival times (headways) of all saturated vehicles are measured at the intersection stop line, a saturated vehicle being one that has had to stop or almost stop in the queue behind the traffic signal. In the headway method, saturation flow is calculated directly as the reciprocal of the average headway of saturated vehicles.

One of the problems in estimating saturation flows from observed headways is accurately defining the saturation flow region of the queue. This problem is usually addressed by one of two procedures. The first and more straightforward of the procedures is to simply plot the average time headways of a queue of vehicles entering an intersection from a stopped position. These plots typically take the form shown in Figure 1. The saturation flow region of the queue can be identified by examining the plot and making a subjective determination of the vehicle storage positions for which departure headways could be assumed to be equal. Formal statistical procedures such as analysis-of-variance and multiple comparisons can be used to examine the departure headways in a more objective manner (e.g., see paper by Stokes, Messer, and Stover elsewhere in this Record).

A second approach to the problem of determining the saturation flow region of the queue involves the use of a formal optimization process such as simple linear regression. The regression models used in this process are typically of the following form:

Texas Transportation Institute, Texas A&M University System, College Station, Tex. 77843.

$$T_c = B_0 + B_1 n_i + e_i \quad (1)$$

where

T_c = elapsed time since the start of green for the vehicle in queue storage position n_i to enter the intersection,
B_0, B_1 = regression coefficients, and
e_i = random error term.

The parameters of interest in Equation 1—B_0 and B_1—represent starting delay and average headway, respectively.

Presented in this paper is an example of the use of simple linear regression to estimate the saturation flows of exclusive double left-turn lanes. The basic procedures utilized are applicable to straight-through movements as well. Perhaps more important, some theoretical considerations and potential bias in saturation flows estimated by simple linear regression are also discussed.

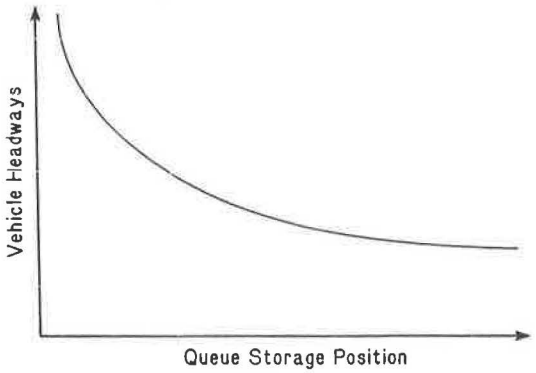


FIGURE 1 Generalization of the relationship between departure headways and queue storage positions.

DATA BASE

The data used in this paper are from a larger study directed at estimating the saturation flows of exclusive double left-turn lanes *(2)*. The data consist of the time (sec) elapsed since the start of the left-turn green phase for successive left-turning vehicles to cross the intersection stop line. The data were collected using time-lapse photography (9 frames/sec) at 14 intersection approaches with exclusive double left-turn lanes in three Texas cities. Table 1 gives a summary of the sample in terms of green phases, left turns, and average queue lengths observed. As indicated by the data in the table more than 3,400 completed left turns were observed at the 14 study sites. This was considered to be a sufficient sample for estimating left-turn saturation flows.

DATA ANALYSIS

The regression analysis involved evaluation of the following model:

$$T_{c(i)} = B_0 + B_1 n_i + e_i \quad (2)$$

where $T_{c(i)}$ is the time (sec) after the start of the left-turn green phase for the rear axle of the vehicle in queue storage position n_i to cross the intersection stop line from Turn Lane i, for i = 1, 2 (see Figure 2).

In Equation 2, the parameter of interest is B_1, where the sample estimate of B_1, denoted by b_1, is an estimate of average departure headway (sec/veh). The parameters of the model given by Equation 2 were estimated by the least-squares method using the regression (REG) procedure of the Statistical Analysis System (SAS) Computer Program Package *(3)*. The

TABLE 1 SUMMARY OF THE SAMPLE

Study Site[a]	No. of Phases Observed	No. of Completed Left-Turns Observed (veh)		Average Queue Length[b] (veh)	
		Lane 1 (inside)	Lane 2 (outside)	Lane 1 (inside)	Lane 2 (outside)
WB US 183 at Burnet Rd., Austin, Tex.	22	175	179	6.5	6.7
EB US 183 at Burnet Rd., Austin, Tex.	12	133	121	8.4	8.6
NB Texas at University, College Station, Tex.	22	171	199	7.0	8.9
WB University at Texas, College Station, Tex.	19	124	118	8.6	9.8
EB University at Texas, College Station, Tex.	13	89	116	6.2	7.5
NB Texas at Jersey, College Station, Tex.	23	231	221	8.0	10.4
EB Jersey at Texas, College Station, Tex.	13	82	95	6.3	7.0
SB Texas at Harvey Rd., College Station, Tex.	15	118	136	6.4	8.9
EB Westheimer at Gessner, Houston, Tex.	23	143	132	6.6	6.6
SB Gessner at Westheimer, Houston, Tex.	22	114	110	6.4	6.6
NB Post Oak Blvd. at San Felipe, Houston, Tex.	15	70	80	4.6	6.7
NB Hillcroft at Westheimer, Houston, Tex.	23	124	134	8.7	8.5
SB Gessner at Bellaire, Houston, Tex.	14	64	41	4.9	2.7
NB Gessner at Bellaire, Houston, Tex.	16	69	69	4.4	5.9
Total	252	1,707	1,751		
Average				6.8	7.6

[a]WB = westbound, EB = eastbound, NB = northbound, and SB = southbound.
[b]Queue Length = number of vehicles stopped at the onset of the left-turn green phase.

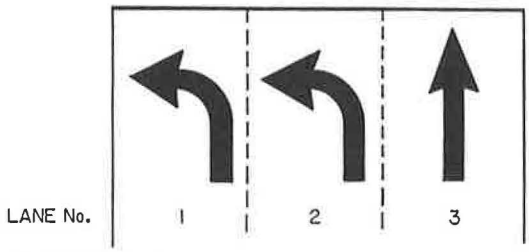


FIGURE 2 Lane numbering scheme

SAS Computer Program Package has been extensively tested and is widely accepted for statistical analyses.

The parameters of the Equation 2 regression model were estimated from the entire sample (i.e., for all vehicles), as well as from subsets of the sample. The subsets of the sample considered included the following regions of the queue:

$1 \leq n_i \leq n_{k(i)}$

.

.

.

$5 \leq n_i \leq n_{k(i)}$

where n_i is the queue storage position for left-turning vehicles in Lane i, for $i = 1, 2$; and $n_{k(i)}$ is the total number of vehicles queued in Lane i at the start of the left-turn green phase.

RESULTS

Table 2 gives the 95 percent confidence intervals for the b_1's for the fitted regression models. Note in this table that the most precise estimates of minimum departure headways were obtained for the region of the queue defined by $2 \leq n_i \leq n_{k(i)}$. Also note that the 95 percent confidence interval for average headways for the region of the queue defined by $3 \leq n_i \leq n_{k(i)}$ contains the point estimate for the region defined by $2 \leq n_i \leq n_{k(i)}$. That is, the estimates obtained from these two regions of the queue differ only in the amount of variation exhibited by the estimates. In addition, the precision in the estimates obtained for the region of the queue defined by $3 \leq n_i \leq n_{k(i)}$ is within 0.1 sec of the estimates obtained for the region given by $2 \leq n_i \leq n_{k(i)}$. Hence, it is suggested by the regression analyses that the saturation flow region of the queue can be defined by the region $3 \leq n_i \leq n_{k(i)}$.

Table 3 gives a summary of the fitted Equation 2 regression models for each of the three cities studied for the saturation flow region of the queue. The saturation flow estimates developed from Equation 2 are given in Table 4.

DISCUSSION OF RESULTS

The concept of saturation flow dictates that when a queue of vehicles is released by a traffic signal, average departure headways gradually decrease until after the first few vehicles, when the headways become uniform. The theory, then, suggests that

TABLE 2 95 PERCENT CONFIDENCE INTERVALS[a] FOR AVERAGE LEFT-TURN DEPARTURE HEADWAYS ESTIMATED FROM EQUATION 2 (SEC/VEH)

	Average Headways (h_L) for Data Sets Shown					
	All	$j \leq n_i \leq n_{k(i)}$, for j =				
Lane	Vehicles	1	2	3	4	5
1	2.2[b]	2.1[b]	2.1[b]	$1.9 \leq \bar{h}_L \leq 2.1$	$1.8 \leq \bar{h}_L \leq 2.0$	$1.8 \leq \bar{h}_L \leq 2.0$
2	2.3[b]	2.2[b]	2.1[b]	$2.0 \leq \bar{h}_L \leq 2.2$	$2.1 \leq \bar{h}_L \leq 2.3$	$2.1 \leq \bar{h}_L \leq 2.3$

[a]Confidence interval = $b_1 \pm 2 \times$ standard error of b_1.
[b]Confidence interval < 0.1 sec.

TABLE 3 SUMMARY OF EQUATION 2 REGRESSION MODELS BY CITY FOR $3 \leq n_i \leq n_{k(i)}$

		b_0 (Starting Delay)		b_1 (Headway)			
City and Lane	Sample Size	Estimate (sec)	Standard Error	Estimate (sec/veh)	Standard Error	Mean Square Error	r^2
Austin							
Lane 1	175	2.2	0.37	1.9	0.06	2.79	0.84
Lane 2	180	1.3	0.31	2.2	0.05	2.09	0.91
College Station							
Lane 1	486	2.5	0.20	1.9	0.04	2.57	0.86
Lane 2	592	2.2	0.26	2.0	0.04	5.07	0.80
Houston							
Lane 1	339	2.4	0.22	1.8	0.05	1.04	0.79
Lane 2	329	2.5	0.24	1.7	0.06	1.33	0.74

TABLE 4 95 PERCENT CONFIDENCE INTERVALS[a] FOR AVERAGE LEFT-TURN SATURATION FLOW BY CITY

Lane	Average Saturation Flow, $\bar{s}_L$ (vphg)		
	Austin	College Station	Houston
1	$1780 \le \bar{s}_L \le 2020$	$1820 \le \bar{s}_L \le 1980$	$1895 \le \bar{s}_L \le 2120$
2	$1565 \le \bar{s}_L \le 1715$	$1730 \le \bar{s}_L \le 1875$	$1980 \le \bar{s}_L \le 2280$

[a]Confidence interval = $1/(\bar{h}_L \pm 2 \times \text{standard error } \bar{h}_L) \times 3{,}600$.

if Equation 2 were to be evaluated by using data sets that successively ignore the first, first and second, . . . , first through fifth vehicles in the queue, the resulting regression lines should tend to become successively more linear. Specifically, it would be expected that the coefficient of determination (r^2) would increase, and that the standard error of b_1 [$S(b_1)$] would decrease as data beginning at successively larger queue storage positions were used to estimate the parameters of the model.

Table 5 gives a summary of the analysis-of-variance tables and headway estimates for Equation 2 as estimated from the total sample and subsets of the sample. An examination of the slopes (i.e., headways) given in Table 5 indicates that they appear to behave in a manner consistent with the concept of saturation flow. That is, as vehicles near the head of the queue are successively deleted from the sample, the slopes tend to decrease, indicating a flattening of the regression lines. However, at first glance the r^2's, and the $S(b_1)$'s to a lesser extent, appear to exhibit somewhat counterintuitive behaviors. For example, there is a tendency for the $S(b_1)$'s to increase and a tendency for the r^2's to decrease as vehicles near the head of the queue are successively eliminated from the data set.

The apparent anomalies in the behavior of r^2 and $S(b_1)$ are probably due to the subsetting of the data, not to some inconsistency with the concept of saturation flow. To avoid misinterpreting the regression models, the nature of the mathematical components of the statistics involved should be examined. In this regard, the following points deserve note.

The value taken by r^2 in a given sample tends to be affected by the range in the observations of the independent variable, the range in the n_i observations in this case. An examination of the components used to calculate r^2, as shown in Equations 3–3e, demonstrates the nature of its seemingly anomalous behavior.

$$r^2 = [\Sigma (T_{c(i)} - \bar{T}_{c(i)})^2 - \Sigma (T_{c(i)} - \hat{T}_{c(i)})^2]/\Sigma (T_{c(i)} - \bar{T}_{c(i)})^2 \quad (3)$$

$$= \Sigma (\hat{T}_{c(i)} - \bar{T}_{c(i)})^2/\Sigma (T_{c(i)} - \bar{T}_{c(i)})^2 \quad (3a)$$

$$= 1 - [\Sigma (T_{c(i)} - \hat{T}_{c(i)})^2/\Sigma (T_{c(i)} - \bar{T}_{c(i)})^2] \quad (3b)$$

where $\bar{T}_{c(i)}$ and $\hat{T}_{c(i)}$ are means and least-squares estimates, respectively. By using a simplified notation, Equations 3–3b can be restated as follows:

$$r^2 = (\text{SST} - \text{SSE})/\text{SST} \quad (3c)$$

$$= \text{SSR}/\text{SST} \quad (3d)$$

$$= 1 - (\text{SSE}/\text{SST}) \quad (3e)$$

TABLE 5 SUMS-OF-SQUARES AND HEADWAY ESTIMATES FOR EQUATION 2

Data Set and Lane	Sample Size	Sums of Squares		Mean Square Error	b_1 (Headway)		r^2
		Error	Total		Estimate (sec/veh)	Standard Error	
All vehicles							
Lane 1	1,707	7,409	74,542	4.4	2.2	0.02	0.90
Lane 2	1,751	10,697	81,841	6.1	2.3	0.02	0.87
$1 \le n_i \le n_{k(i)}$							
Lane 1	1,500	3,028	35,077	2.0	2.1	0.02	0.91
Lane 2	1,598	4,831	49,182	3.0	2.2	0.02	0.90
$2 \le n_i \le n_{k(i)}$							
Lane 1	1,251	2,606	23,143	2.1	2.0	0.02	0.89
Lane 2	1,347	4,541	34,615	3.4	2.1	0.02	0.87
$3 \le n_i \le n_{k(i)}$							
Lane 1	1,000	2,212	15,431	2.2	2.0	0.03	0.86
Lane 2	1,101	4,193	24,669	3.8	2.1	0.03	0.83
$4 \le n_i \le n_{k(i)}$							
Lane 1	752	1,763	10,054	2.4	1.9	0.03	0.82
Lane 2	859	3,792	17,513	4.4	2.2	0.04	0.78
$5 \le n_i \le n_{k(i)}$							
Lane 1	509	1,297	6,318	2.6	1.9	0.04	0.79
Lane 2	623	3,356	12,074	5.4	2.2	0.05	0.72

where

SST = total sums of squares,
SSE = error sums of squares, and
SSR = regression (or model) sums of squares.

In general, SSE is not affected systematically by the spacing of the n_i's. However, the wider the spacing of the n_i's in the sample, the greater the spread of the observed $T_{c(i)}$'s around $T_{c(i)}$ will tend to be, and hence the greater will be SST. Consequently, the wider the n_i's are spaced, the higher r^2 will tend to be *(5)*. Thus, the observed reductions in the r^2's are due to the narrowing of the range in the n_i's associated with the successive elimination of vehicles near the head of the queue.

A similar explanation applies to the behavior of $S(b_1)$ across the subsamples. $S(b_1)$ is given by

$$S(b_1) = [\text{MSE}/\Sigma(n_i - \bar{n}_i)^2]^{1/2} \quad (4)$$

where MSE is the mean square error as given by SSE divided by the degrees of freedom for error.

As the sample is truncated (by successively eliminating vehicles at the head of the queue), the $\Sigma\ (n_i - \bar{n}_i)^2$ term in Equation 4 must decrease. Consequently, as long as the MSEs remain fairly stable across the subsamples, $S(b_1)$ must increase as one moves across the subsamples. An examination of the data in Table 5 suggests that for saturated vehicles the MSEs for each lane are fairly uniform across the subsamples. Consequently, the increases in $S(b_1)$ would appear to be entirely logical.

The point of the preceding discussion is that a simple comparison of the r^2's and $S(b_1)$'s for the queue segments considered may not be an efficient procedure for identifying the saturation flow region of the queue.

It also appears that at least one of the basic assumptions of regression analysis was not satisfied. Specifically, the variance of the error term in the regression models does not appear to be constant across the sample. Figures 3 and 4, which show plots of the residuals against the queue storage positions by lane for the region of the queue defined by $1 \leq n_i \leq n_{k(i)}$, illustrate the problem. The patterns of the plots suggest that the error vari-

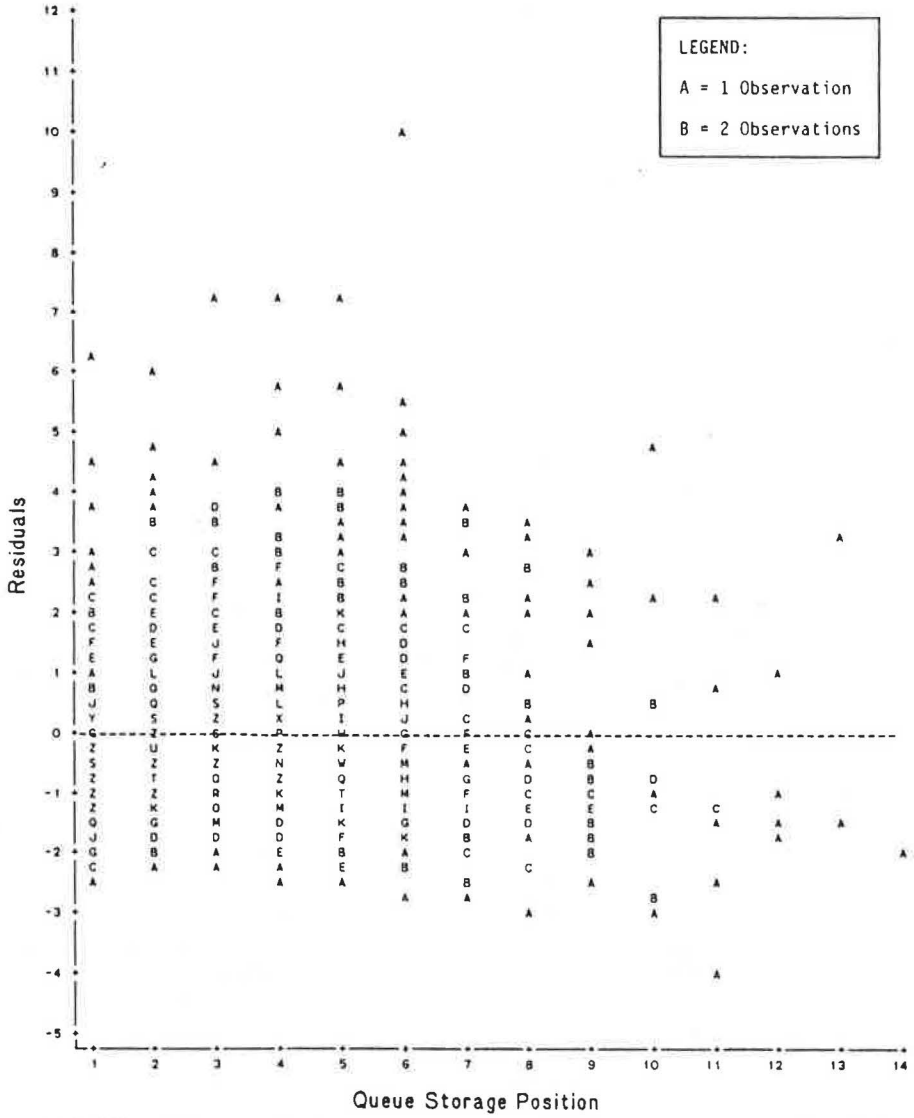


FIGURE 3 Plot of residuals against queue storage positions for Equation 2, for all approaches, for $1 \leq n_i \leq n_{k(i)}$: Lane 1.

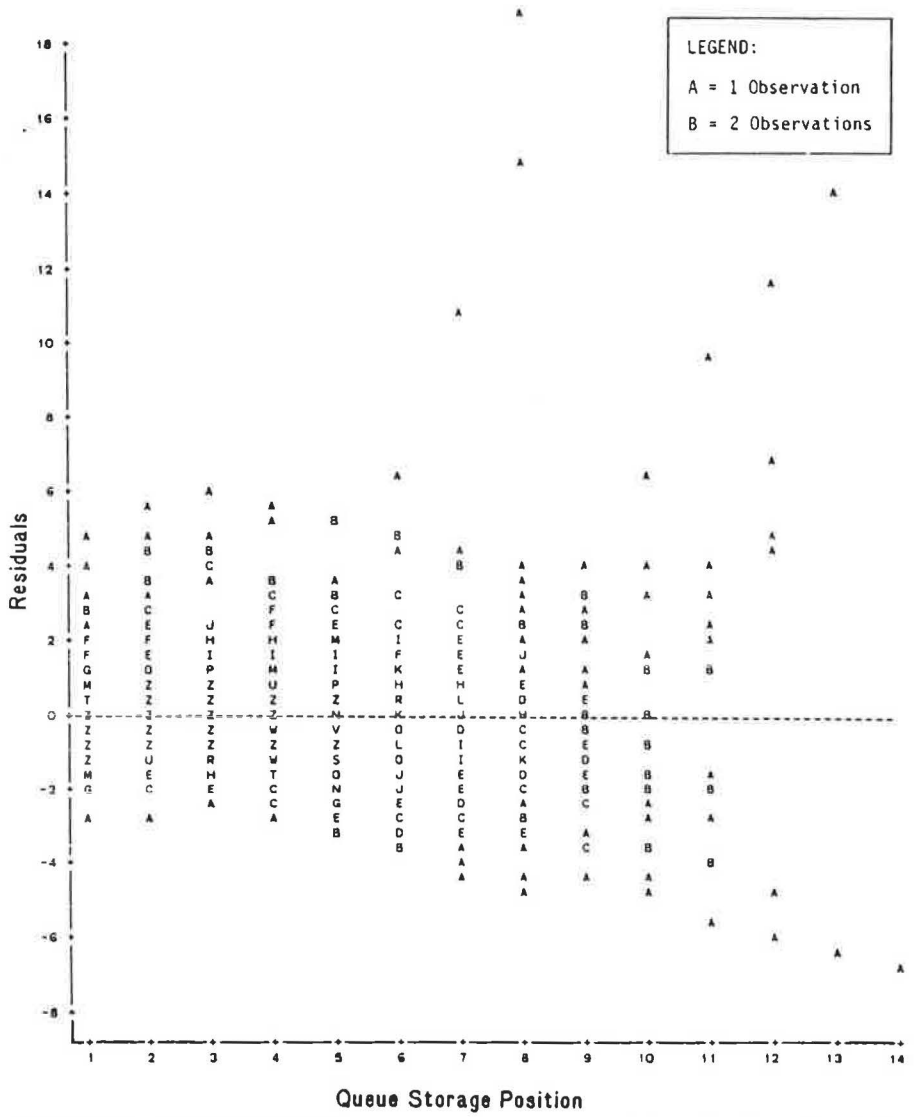


FIGURE 4 Plot of residuals against queue storage positions for Equation 2, for all approaches, for $1 \leq n_1 \leq n_{k(i)}$: Lane 2.

ance is a function of queue storage position. The patterns depicted for the two lanes are intriguing. For example, in the case of Lane 1 (Figure 3), the error variance tends to increase with queue storage position up to about the sixth position, where it begins to fall off and then stabilize. Note in Figure 4 that Lane 2 (the outside lane) exhibits the opposite trend; that is, the error variance is fairly constant until about the sixth storage position, where it begins to increase.

In addition to the subjective analysis of residuals just described, the constancy of the error variance was evaluated by using a formal statistical procedure. The procedure used was to

1. Array the observations by queue storage position and lane,
2. Divide the total observations into two equal data sets for each lane,
3. Fit separate regression functions to each half of the total observations,
4. Calculate the MSE for each, and
5. Test for equality of the error variances by the F-test.

The resulting variance ratios were significant at the 5 percent level.

Thus, evaluation of the error variance suggests that the regression estimates of the departure headways may be biased. Table 3 presents data on the nature of the suspected bias. Note that, with the exception of Lane 2 for the Austin sites, the estimated headways (i.e., the slopes) are all less than or equal to 2.0 sec. The regression estimates suggest average left-turn flow rates in excess of 1,800 vehicles per hour of green time per lane (vphgl) for the majority of the study sites within each city. Relative to generally accepted straight-through flow rates of 1,700-1,800 vphgl, the regression models appear to have underestimated average left-turn departure headways.

SUMMARY

In using least-squares regression techniques in model-building exercises, there are a number of potential problem areas that are frequently overlooked by researchers using this method. For example, it is important that the researcher validate the

distributional assumptions on the model errors. The standard assumptions are that the dependent variable is normally distributed and that the errors are independent and have homogeneous variances.

In this study, the possible consequences of overlooking the distributional assumptions about the error variances have been examined. The results of this limited study suggest that nonconstancy of the error variances (heteroscedasticity) may result in regression models that substantially underestimate left-turn departure headways. It is hoped that the discussion relating to model estimation and validation will encourage others to address these basic issues in the literature.

ACKNOWLEDGMENTS

The authors wish to thank the following individuals for their contributions to the study: D. L. Pugh, J. deJong, D. A. Maxwell, L. J. Ringer, and A. M. Elmquist, all with Texas A&M University; D. W. Hall, City of Austin; J. R. Black, City of College Station; and W. E. Hensch, City of Houston. The authors remain solely responsible for the contents of this paper.

REFERENCES

1. F. V. Webster and B. M. Cobbe. *Traffic Signals.* Her Majesty's Stationery Office, London, England, 1966.
2. R. W. Stokes. *Saturation Flows of Exclusive Double Left-Turn Lanes.* Ph.D. dissertation. Texas A&M University, College Station, 1984.
3. *SAS User's Guide: Statistics.* SAS Institute, Inc., Cary, N.C., 1982.
4. J. Neter, W. Wasserman, and M. H. Kutner. *Applied Linear Regression Models.* Richard D. Irwin, Inc., Homewood, Ill., 1983.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Freeway Weaving Sections: Comparison and Refinement of Design and Operations Analysis Procedures

JOSEPH FAZIO AND NAGUI M. ROUPHAIL

Weaving sections represent the common right-of-way that occurs when two or more crossing freeway traffic streams are traveling in the same general direction. In conjunction with the development of the 1985 *Highway Capacity Manual* (HCM), several procedures have evolved for the purpose of updating, revising, and replacing the 1965 HCM procedure for design and operations analysis of freeway weaving sections. The objectives of this paper are twofold: to present and review the latest three weaving procedures available to highway and traffic engineers, and to propose specific refinements to a simple weaving section procedure to account for the lane distribution of traffic upstream of the weaving section. These adjustments primarily involve the development of a lane-shift variable, which represents the average amount of peak-period passenger car lane shifts occurring under a given geometric configuration and prevailing traffic volumes. Statistical testing of the refined procedure against the three procedures at more than 50 sites nationwide indicated that the proposed procedure tends to predict observed average running weaving and nonweaving speeds more closely than do the other procedures in most cases.

A weaving section represents the physical space along a freeway where two (simple weaving) or more (multiple weaving) traffic streams traveling in the same general direction cross each other. Four basic movements are serviced in a simple weaving section, two weaving and two nonweaving (outer flows), as indicated in Figure 1*a*. Weaving traffic originating from the freeway mainline is denoted V_2 and nonweaving traffic is denoted V_1. Weaving traffic originating from the minor approach or entrance ramp is denoted V_3 and nonweaving traffic is denoted V_4. The length of a weaving section *(L)* and the number of lanes *(N)* are the two design parameters that dictate the mode of traffic operation to be expected, as illustrated in Figure 1*b*. (Note in this figure that N_b is the basic

Urban Transportation Center and CEMM Department, University of Illinois, Chicago, Ill. 60680.

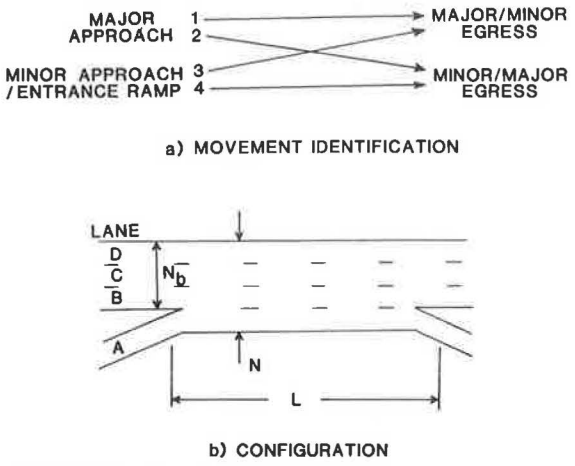


FIGURE 1 Simple freeway weaving section: movement identification and configuration.

number of lanes on a major approach to a weaving section. The number of lanes entering the weaving section from the minor approach or entrance ramp is denoted N_r.) As L or N decreases, drivers must execute their lane changes in a relatively short space, thus resulting in a general decrease in speed and level of service (LOS) for all traffic. In addition, lane changes originating from the outer lanes (i.e., median lane on the main approach and shoulder lane on the minor approach) will tend to create increased disruption to traffic operations in the weaving section compared with the situation resulting from a presegregated traffic stream, when weaving traffic is essentially confined to the boundary lanes between the two traffic streams.

Design procedures for weaving sections are aimed at determining the minimum length and number of lanes in the weaving section needed to meet a prespecified LOS in the analysis period. Operations analysis involves the determination of LOS and average running speeds for an existing weaving section.

BACKGROUND

One of the earliest weaving procedures for the design and operations analysis of weaving sections for the nation's first freeways appeared in the 1950 *Highway Capacity Manual (1)*. The development of this procedure was based on field data collected at six weaving sites. In 1957, the 1950 HCM procedure was updated with additional field data *(2)*. A major data collection effort was undertaken by the Bureau of Public Roads in 1963, which resulted in a new weaving analysis procedure in the 1965 HCM *(3)*. This method has been widely used during the past two decades and constitutes the current state of practice for design and analysis of weaving sections.

In an effort to keep abreast of changes in traffic composition and characteristics that took place since the Bureau of Public Roads data were collected, Project 3-15, Weaving Area Operations Study, was initiated in 1969 through the National Cooperative Highway Research Program *(4)*. This study, conducted by the Polytechnic Institute of New York (PINY), included field data collection at 17 northeastern sites. The results of Project 3-15 ultimately led to an interim weaving procedure published in 1980 *(5)*. In an independent effort, a nomographic weaving procedure initially published in 1979 *(6)*, was also included in the *Interim Materials for Highway Capacity;* this procedure was further modified in 1984 [see Leisch *(7)*]. Recently, the FHWA, U.S. Department of Transportation, initiated a study to evaluate the two procedures; the study resulted in yet another procedure for analyzing weaving sections [see JHK *(8)*]. Based on the conclusions of the JHK report, the PINY procedure was revised and eventually adopted as the weaving procedure for the 1985 HCM *(9)*. Since that time, both the Leisch and JHK procedures have undergone further revisions.

The procedures reviewed in this paper are the 1985 HCM (PINY) procedure, the revised JHK procedure (based on weaving study memoranda by W. Reilly and P Johnson, JHK and Associates, November 1984), and revised Leisch procedure (based on information letter from J. Leisch of J. Leisch and Associates, February 1985).

COMPARISON OF WEAVING ANALYSIS PROCEDURES

Tables 1 and 2 give summaries of the input requirements and output obtained for each of the three procedures. Of the three weaving procedures, the JHK procedure is the simplest to use.

TABLE 1 COMPARISON OF INPUT REQUIREMENTS FOR THREE WEAVING PROCEDURES

Method	Configuration	N_b	N	L	V	V_W	V_{W2}	V_4
JHK			X	X	X	X		X
Leisch	X	X	X	X	X	X	X	
1985 HCM	X		X	X	X	X	X	

Note: N = number of lanes within weaving section, L = length of the weaving section measured from the point at which the entrance gore is 2 ft wide to the point at which the exit gore is 12 ft wide, V = total volume of traffic in the weaving section = $V_1 + V_2 + V_3 + V_4$, V_1 = volume of nonweaving traffic stream originating from the major approach to the weaving section, V_2 = volume of weaving traffic stream originating from the major approach to the weaving section, V_3 = volume of weaving traffic stream originating from the entrance ramp or minor approach to the weaving section, V_4 = volume of nonweaving traffic stream originating from the entrance ramp or minor approach to the weaving section, V_W = volume of weaving traffic in the weaving section = $V_2 + V_3$, and V_{W2} = volume of smaller of the two weaving traffic streams [min (V_2, V_3)].

In essence, this procedure utilizes two equations for average running speeds, one for weaving, and the other for nonweaving traffic, in Equations 1 and 2:

$$S_W = 15 + \left[50/\left(1 + \{2{,}000[1 + (V_4/V)]^{2.7}\,[1 + (V_W/V)]^{0.9} \times [V/(QN)]^{0.6}/L^{1.8}\}\right)\right] \quad (1)$$

$$S_{NW} = 15 + \left[50/\left(1 + \{100\,[1 + (V_4/V)]^{5.4} \times [1 + (V_W/V)]^{1.8}\,[V/(QN)]^{0.9}/L^{1.8}\}\right)\right] \quad (2)$$

where Q is the heavy vehicle factor, and the other variables are as defined in Tables 1 and 2.

To use the JHK equations, hourly volumes must be adjusted to passenger car equivalents via the heavy vehicle factor *(Q)*.

TABLE 2 COMPARISON OF OUTPUT GENERATED BY THREE WEAVING PROCEDURES

Method	S_W	S_{NW}	S	LOS_W	LOS_{NW}	LOS_T	N_W	SF	Operation Mode
JHK	X	X		X	X				
Leisch	X		X	X		X		X	
1985 HCM	X	X		X	X		X		X

Note: S_W = average running speed of weaving traffic in the weaving section (mph), S_{NW} = average running speed of nonweaving traffic in the weaving section (mph), S = average running speed of all traffic in the weaving section (mph), LOS_W = level of service for weaving traffic, LOS_{NW} = level of service for nonweaving traffic, LOS_T = level of service for all traffic within the weaving section, N_W = theoretical number of lanes used by weaving traffic in the weaving section, and SF = service flow (pcphpl).

After the average running weaving and nonweaving speeds are calculated from Equations 1 and 2, weaving and nonweaving levels of service are read out from appropriate tables.

The Leisch procedure is nomograph-oriented, as shown in Figure 2. [Note in Figure 2 that R is the weaving ratio (V_{w2}/V_w). All other variables are defined elsewhere in the paper.] Two nomographs are used for one-sided weaving sections and two for two-sided sections. Configuration is accounted for in the procedure by *(a)* categorizing the weaving section as one sided or two sided, *(b)* specifying the presence or absence of lane balance (lane balance occurs when the combined number of exit lanes on the freeway and ramp is one more than the number of lanes on the freeway within the weaving section), and *(c)* providing for an approximate reduction in traffic speeds when the section configuration is concomitant with an excessive amount of lane shifts. Peak-hour factor values are built into the procedure, thus requiring no adjustments for peak-hour flow, except for vehicle composition. The procedure derives the average running speed for weaving traffic and overall average running speed within the weaving section. Also determined by the procedure are service flow [service volume in passenger cars per hour per lane (pcphpl)], weaving intensity factor *(k)*, LOS_W, and LOS_T, as defined in Table 2.

The 1985 HCM weaving procedure uses the following equation to estimate average running weaving and nonweaving speeds.

$$S_W \text{ or } S_{NW} = 15 + \left[50/\left(1 + \{a\,[1 + (V_W/V)^b \times (V/N)^c/L^d\}\right)\right] \quad (3)$$

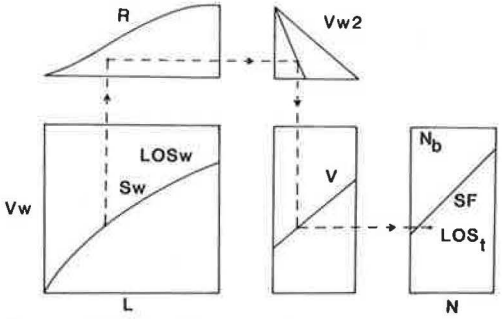


FIGURE 2 Leisch procedure: nomograph outline.

where a, b, c and d are calibration constants based on section configuration and type of operation. The method categorizes weaving sections into Types A, B, and C as a function of the minimum number of lane shifts performed by a driver in each of the two weaving traffic streams, as indicated in the HCM Weaving Chapter (*9*, Table 4-1).

A key element in the HCM procedure is whether traffic operation is constrained or unconstrained. This is determined by comparing the number of lanes required for unconstrained operation, N_W, with the theoretical value N_W (Max) [see HCM Table 4-4 *(9)*]. Weaving and nonweaving speeds are then determined from Equation 3 based on configuration type and operation mode for the section under consideration. Finally, two levels of service are determined separately for weaving and nonweaving traffic [see HCM Table 4-6 *(9)*].

Another important aspect of all procedures is the range of operating conditions in which a solution can be found. Inherent limitations in the procedures include section geometry [i.e., maximum values for L, N, or N_b, section capacity V_W, SF, weaving frequencies VR (where VR is volume ratio = V_W/V), R (where R is weaving ratio = V_{W2}/V_W), and running speeds S_W, S_{NW}]. A comparison of the three procedures in that respect is given in Table 3.

LANE SHIFT CONCEPT

A noticeable difference between the JHK and HCM procedures is that the latter introduces the configuration of the weaving section into the speed equation *(9)*. The parameters a, b, c, and d in the HCM method are determined in part from the minimum number of lane shifts required by a driver in each of the weaving streams (as indicated in HCM Table 4-1), thus implying that weaving traffic is completely segregated on entering the weaving section.

Field measurements collected recently indicate that weaving traffic is not fully segregated on entering the weaving section (based on information letter from Eric Ruehr, JHK and Associates, October 1984). For $N_B = 2$, it was found that an average of 93.4 percent of Movement 2 traffic entered by way of Lane B (Figures 1*a* and 1*b*), while 6.6 percent entered via Lane C. For $N_B \geq 3$, only 90.5 percent of Movement 2 traffic entered the weaving section by way of Lane B, with almost 10 percent of all traffic arriving in Lanes C and D. A negligible percentage of vehicles arrive in the outer lanes E, F, and so forth. A summary

TABLE 3 COMPARISON OF PROCEDURE LIMITATIONS

Procedure	Parameter	Limitation	Comments
Leisch	S_{Wi}	~ ≤ 55 mph	Initial average running speed of weaving traffic out of realm of weaving
	S_{Wf}	~ ≤ 55 mph	Same as above, for final weaving speed
	SF	~ 2000 pcphpl	Service flow beyond nomograph boundary
JHK	S_W	15 mph < S_W < 65 mph	Outside the realm
	S_{NW}	15 mph < S_{NW} < 65 mph	of weaving
	L	≤ 4,000 ft	
1985 HCM	V_W	Type A, 1,800 pcph Type B, 3,000 pcph Type C, 3,000 pcph	Weaving section capacity
	V/N	1,900 pcph (A, B, C)	Lane capacity
	VR	Type A: N = 2, 1.00 N = 3, 0.45 N = 4, 0.35 N = 5, 0.22 Type B: 0.80 Type C: 0.50	Volume ratio limits
	R	Type A: 0.50 Type B: 0.50 Type C: 0.40	Weaving ratio limits
	L	Type A: 2,000 ft Types B and C: 2,500 ft	Length out of realm of weaving
	S_W	15 < S_W < 65	
	S_{NW}	15 < S_{NW} < 65	

of the observed distribution of Movement 2 vehicles is given in Table 4.

As a logical extension of the results just given, an index was developed that takes into account the interaction of the following

- Weaving volumes V_2 and V_3,
- Distribution of V_2 and V_3 across lanes, and
- The minimum number of lane shifts by lane of entry.

A lane shift multiplier has been developed that represents the minimum number of lane shifts that must be executed by the driver of a weaving vehicle from his lane of origin to the closest destination lane. This parameter can be determined directly from a sketch of the existing or proposed weaving section. Two examples, with balanced and imbalanced sections, which demonstrate the computation of the lane shift multipliers A, B, C, and D are shown in Figures 3*a* and 3*b*, respectively. Note the following in Figure 3:

A = lane shift multiplier for entering lane A (LS/veh);
B = lane shift multiplier for entering lane B (LS/veh);
C = lane shift multiplier for entering lane C (LS/veh);
D = lane shift multiplier for entering lane D (LS/veh);

From the previous analysis, the total number of peak-hour lane shifts performed in the weaving section can be calculated. When adjusted for variations in vehicle and driver population, peak-hour factor (PHF), and lateral clearances, the resulting index, termed passenger car lane shifts per hour (pcLSph) provides a means for integrating several operating parameters of the weaving section into a single variable. The index also avoids the artificial designation of weaving sections into Type A, B, and so forth; rather, it provides the traffic engineer with a single numeric value that is indicative of the level of maneuvering difficulty encountered by all drivers in the weaving section.

TABLE 4 OBSERVED LANE DISTRIBUTION OF TRAFFIC UPSTREAM OF WEAVING SECTIONS

JHK Site No.	Percent Movement 2 Traffic in Indicated Lane[a]				
	N_b = 2		N_b = 3		
	B	C	B	C	D
1			93.1	6.9	0.0
2	97.0	3.0			
3	89.7	10.3			
4[b]			91.1	8.4	0.5
			95.1	4.5	0.0
			88.3	9.2	2.5
5[b]			84.3	14.4	1.3
			92.2	6.8	1.0
			88.8	9.4	1.8
Avg	93.4	6.6	90.5	8.5	1.0

Source: Information letter from Eric Ruehr, JHK and Associates, October 1984.

[a]See Figure 1 for lane designation.
[b]Multiple observations per site.

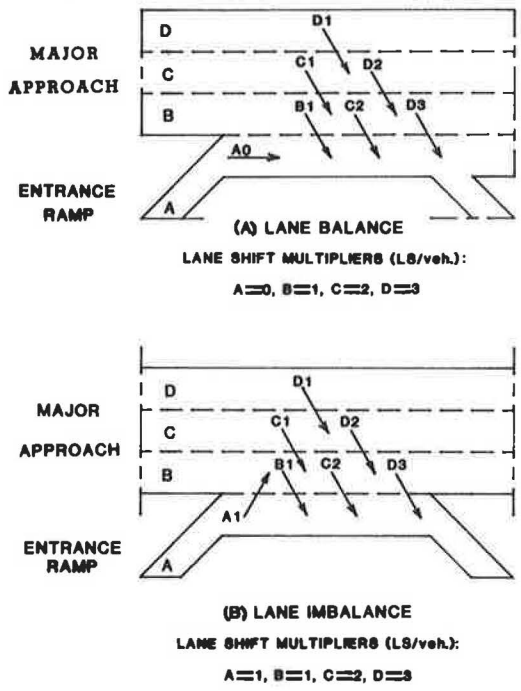


FIGURE 3 Examples of determining lane shift multipliers.

Equations for determining the lane shift index are given in Table 5 for different configurations of weaving sections. Note the following in Table 5:

- LS = average number of lane shifts performed by the drivers of weaving vehicles = $LS_2 + LS_3$ [passenger car lane shifts per hour (pcLSph)];
- LS_2 = average number of lane shifts performed by the drivers of the Movement 2 vehicles (pcLSph); and
- LS_3 = average amount of lane shifts performed by the drivers of the Movement 3 vehicles (pcLSph).

Initial testing of the lane shift index consisted of a correlation analysis between the index and average running weaving and nonweaving speeds observed at six sites comprising a total of 12 cases. Examination of the data indicated an inverse relationship between the two parameters, as suspected. Further testing pertaining to the form of the variable indicated that LS/L and LS_3/V correlated well with average running weaving speeds, whereas LS_3/LS correlated well with average running nonweaving speeds. The resulting speed models, which represent an extension of the JHK and 1985 HCM models, are expressed by the following equations:

$$S_W = 15 + \left\{50/\left[\left(1 + \left\{\left[1 + (V_3 + V_4)/V\right]^{3.045}\ (V/N)^{0.605} \times (LS/L)^{0.902}\right\}\right)/75.959\ \left[1 + (LS_3/V)\right]^{3.94}\right]\right\} \quad (4)$$

and

$$S_{NW} = 15 + \left\{50/\left[1 + \left(\left\{\left[1 + (V_4/V)\right]^{5.08}\ \left[1 + (V_W/V)\right]^{2.019} \times (V/N)^{1.523}\right\}/\left\{60.995\ \left[1 + (LS_3/LS)\right]^{0.916}\ L^{1.07}\right\}\right)\right]\right\} \quad (5)$$

It should be noted that the calibration data set for the speed models consisted of 56 cases, including 35 sites from the Bureau of Public Roads study *(3)* and 6 from the JHK study [Reilly and Johnson *(8)*, and weaving study memoranda by Reilly and Johnson, JHK and Associates, November 1984].

PRELIMINARY MODEL EVALUATION

To determine that the proposed speed models are an improvement over other procedures available, a comparison with the JHK models was performed. This task required the recalibration of the JHK models using the set of 56 data points mentioned in the previous section. After both models were calibrated, they were utilized to predict average weaving and nonweaving speeds in 11 validation cases [Reilly and Johnson *(8)*, and weaving study memoranda by Reilly and Johnson, JHK and Associates, November 1984]. A simple regression model between field and predicted average running speeds was developed to test the predictive power of each model. The results are presented in Table 6. As can be observed, the proposed model exhibited higher correlations with observed weaving speeds compared with the recalibrated JHK model. Both models exhibited modest correlations with average nonweaving speeds; the JHK procedure had a slight edge over the proposed model.

COMPARISON OF THE FOUR PROCEDURES

The proposed model has been expanded to a step-by-step procedure for the design and operations analysis of simple weaving sections. Details of the procedure may be found elsewhere *(10)*. In addition, an interactive, microcomputer-based program has been developed that performs all of the calculations necessary to carry out the four procedures described in this paper *(11)*. A set of 67 cases representing the full data base available to the research staff was processed

TABLE 5 LANE SHIFT INDEX EQUATIONS

N_b	LS_2	LS_3
1	$V_2B/(PHF * f_{HV} * f_W * fp)$	$V_3A/(PHF*f_{HV} * f_W * f_p)$
2	$(0.934V_2B + 0.066V_2C)/(PHF * f_{HV} * f_W * fp)$	$V_3A/(PHF*f_{HV}*f_W*fp)$
≥ 3	$(0.905V_2B + 0.085V_2C + 0.010V_2D)/(PHF * f_{HV} * f_W * fp)$	$V_3A/(PHF*f_{HV} * f_W * f_p)$

Note: f_{HV} = heavy vehicle adjustment factor; f_W = lateral clearance adjustment factor; fp = driver population adjustment factor; and all other variables are as defined in the text or previous tables.

TABLE 6 FIELD VERSUS PREDICTED SPEEDS: RESULTS OF TWO MODELS

Parameter	Data Set Type: Calibration[a] JHK[c]	Calibration[a] Proposed Model	Validation[b] JHK	Validation[b] Proposed Model
Weaving speeds				
r^d	0.62	0.74	0.56	0.65
Slope	0.90	0.89	0.62	0.63
Intercept	3.50	3.90	12.20	13.30
Nonweaving speeds				
r	0.53	0.55	0.46	0.40
Slope	0.80	0.79	0.82	0.66
Intercept	7.20	7.70	8.80	16.60

[a]56 cases.
[b]11 cases.
[c]Recalibrated model.
[d]Correlation coefficient.

through the microcomputer program [Reilly and Johnson *(8)*, and weaving study memoranda by Reilly and Johnson, JHK and Associates, November 1984]. A detailed breakdown of the study sites is given in Table 7.

A comparative summary of the results obtained is given in Table 8. As can be observed, the proposed procedure produced the highest correlation with field weaving ($r = 0.72$) and nonweaving speeds ($r = 0.53$). In contrast, the HCM procedure ranked last in correlation with weaving ($r = 0.56$) as well as nonweaving speeds ($r = 0.31$). The JHK and Leisch procedures yielded almost identical correlations for both speeds. Furthermore, the proposed procedure produced the lowest intercept (for perfect correlation, intercept approaches zero) and second highest slope (for perfect correlation, slope approaches unity) compared with the other three procedures.

An assessment of the applicability of each procedure, is presented in Table 9. In this table, the number of sites for which a solution could not be found as a result of inherent operational limitations in each procedure (given in Table 3) is listed for each procedure. Of the 67 cases making up the data base, the 1985 HCM procedure could only be applied to 39 cases. It appears that the limitation on weaving section capacity of 1,800 pcph for Type A configuration and 3,000 pcph for Types B and C resulted in the rejection of many sites in the data base. It is interesting to note that a solution that disregards these limitations produces an estimate of speeds that is in close agreement with some of the field observations.

To confirm that the majority of invalid cases are indeed reflective of the HCM weaving capacity limitations and not due to site anomalies such as excessive length and or number of lanes, the frequencies of all Type A sections were compared with the frequencies of those invalid Type A configuration cases in which V_W exceeded 1,800 pcph. The comparisons were made with respect to length (Figure 4) and number of lanes within the weaving sections (Figure 5). In both instances, the invalid case frequencies closely parallel all Type A frequencies; in other words, the frequencies of invalid cases did not progressively increase as length or number of lanes increased. Similar patterns were observed were observed when the frequencies of Types B and C configurations were compared. No such limitation problems were encountered with the other pro-

TABLE 7 GEOGRAPHICAL DISTRIBUTION OF STUDY SITES

Location	No. of Cases[a]	Method(s) That Used Case(s) for Calibration[b]	No. of Sites
Arlington, Virginia	2 A	L, H, P	2
	1 A	L	
Atlanta, Georgia	1 B	J, P	2
	1 B	L, J	
Boston, Massachusetts	1 A		1
Chicago, Illinois	13 A	L, H, P	14
	1 B	L, J	
	1 A	L	
Gowanus Expressway, New York	1 A		1
Long Island, New York	4 A	L, H, P	3
Los Angeles, California	6 A	L, H, P	6
New York, New York	8 A	L, H, P	5
	1 A		
San Diego, California	2 A		2
San Francisco, California	8 A	L, H, P	9
	9 B	J, P	
Washington, D.C.	3 A	L, H, P	5
	2 B	J, P	
White Plains, New York	1 A		1
Yonkers, New York	1 A		1
Total	67		52

Source: Reilly and Johnson *(8)* and weaving study memoranda by Reilly and Johnson, JHK and Associates, November 1984.

[a]A = pre-1970 data and B = post-1970 data.
[b]L = Leisch weaving procedure, H = 1985 HCM weaving procedure, J = JHK weaving procedure, and P = proposed weaving procedure.

TABLE 8 SUMMARY EVALUATION OF FIELD VERSUS METHOD AVERAGE RUNNING SPEEDS

	Weaving Speed (mph)						Nonweaving Speed (mph)					
Method	*r*	Slope	Intercept (mph)	Standard Deviation Field (mph)	Standard Deviation Method (mph)	Absolute Mean Difference (mph)	*r*	Slope	Intercept (mph)	Standard Deviation Field (mph)	Standard Deviation Method (mph)	Absolute Mean Difference (mph)
Proposed	0.72	0.84	5.4	9.7	8.4	5.4	0.53	0.79	8.2	11.6	7.8	7.9
JHK	0.62	0.65	10.4	10.0	9.6	5.9	0.48	0.60	13.6	11.8	9.4	8.6
Leisch[a]	0.63	0.85	7.7	9.6	7.1	5.7	0.48	0.72	10.5	11.4	7.7	8.4
1985 HCM	0.56	0.82	7.3	7.9	5.3	4.7	0.31	0.42	24.7	11.0	8.0	8.7

[a]Nonweaving speeds in this procedure cannot be directly estimated. Observed nonweaving speeds are correlated with overall speeds, as recommended by the author.

TABLE 9 SUMMARY EVALUATION OF METHOD APPLICABILITY

		No. of	No. of Invalid Cases Due To[a]								
Method	No of Cases	Valid Cases	$V_W >$ 1,800 pcph	$SF >$ 2,000 pcphpl	$VR > 0.22$	$V_W >$ 3,000 pcph	Out of Realm[b]	$R > 0.4$	$L >$ 2,500 ft	4,000 ft	$VR > 0.5$
Leisch	67	65	–	1	–	–	1	–	–	–	–
JHK	67	63	–	–	–	–	–	–	–	4	–
1985 HCM[c]	67	39	9	–	1	9	–	1	6	–	2
Proposed[c]	67	67	–	–	–	–	–	–	–	–	–

[a]In which at least one constraint was violated.
[b]Considered to be beyond the realm of weaving.
[c]Based on 5-min peak flow rates.

cedures: the JHK procedure was applicable in 63 cases, the Leisch procedure in 65 cases, and the proposed procedure in all 67 cases making up the data base.

FINAL NOTE ON THE ANALYSIS PERIOD

Although the decision to calibrate speed models based on hourly or peak flow rates is highly controversial, a simple rule exists when the final models are to be tested: follow the appropriate input requirements stipulated by the method.

In this study, hourly volumes were not adjusted for peak periods in either the Leisch or JHK procedure; the former procedure automatically performs PHF adjustments in the nomographs, whereas the latter does not consider any automatic peak-period adjustments. The proposed procedure and HCM procedure require peak-period adjustments for 5- and 15-min peak flow rates, respectively.

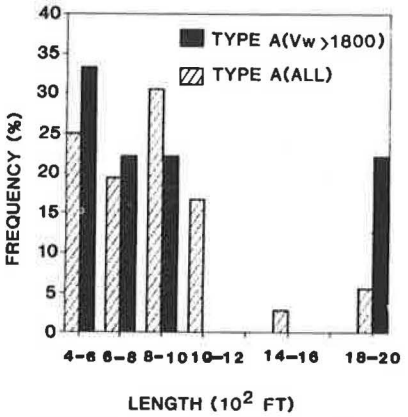


FIGURE 4 Distribution of section length: rejected HCM cases ($V_W >$ 1,800 pcph) versus all Type A.

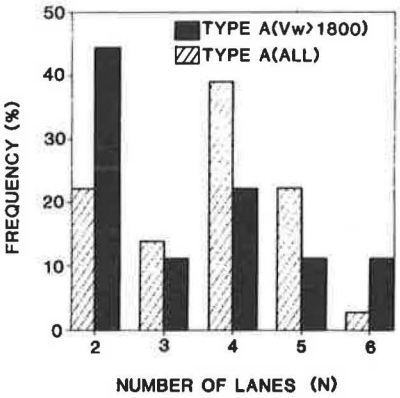


FIGURE 5 Distribution of No. of lanes: rejected HCM cases ($V_W >$ 1,800 pcph) versus all Type A.

However, due to the lack of 15-min data, both procedures were tested based on 5-min peak flow rates. Although it is anticipated that some cases may no longer be invalid under the 15-min assumption, results indicated that the majority of cases that were rejected under the original test ($V_W > 1{,}800$ pcph for Type A, $V_W > 3{,}000$ pcph for Types B and C) remain invalid even when a PHF of 1.0 is assumed (12 out of 18 cases). The true number of rejected cases will probably range between 12 and 18. Even in the absence of such information, it is evident from Equation 3 that PHF (in the term *V/N*) does not signifi-

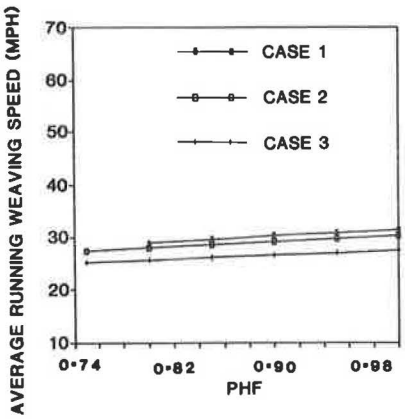


FIGURE 6 Sensitivity of HCM average weaving speed to peak-hour factor for $N = 2$.

cantly affect the operation of the weaving section. Results for three cases from the data base plotted in Figures 6 and 7 for N = 2 (the most critical value for PHF) indicate little variation in predicted weaving and nonweaving speeds in response to variations in the peak-hour factor.

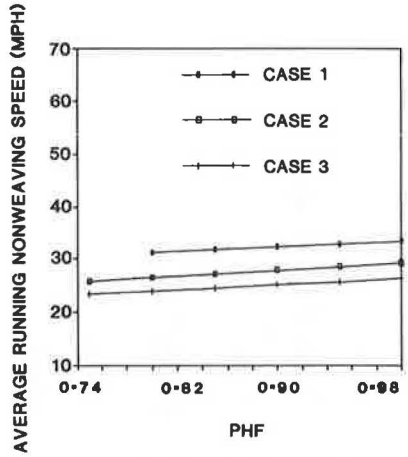


FIGURE 7 Sensitivity of HCM average nonweaving speed to peak-hour factor for $N = 2$.

CONCLUSIONS AND RECOMMENDATIONS

This study was designed to investigate several freeway weaving analysis procedures that were contemplated for the 1985 HCM. It has resulted in the development of a weaving procedure that is superior in predicting weaving and nonweaving speeds compared with existing procedures. The following conclusions are offered.

- The total number of lane shifts required by drivers in weaving sections affects both weaving and nonweaving speeds. Negative correlations between speeds and lane shifts were observed in the field.
- The inclusion of lane shift as an independent variable in average running weaving and nonweaving speed models enhanced the predictive ability of the models considerably. When compared with the latest procedures developed by JHK, Leisch, and the 1985 HCM (PINY), the proposed models yielded the highest correlations with field weaving and nonweaving speeds.
- The 1985 HCM procedure appears to be severely limited in its application; more than 41 percent of all cases analyzed in this study did not meet the constraints stipulated by the method. The majority of the cases failed to satisfy the constraints on weaving section capacity; a majority of these cases would still have been rejected even if hourly rates had been used instead of peak 5-min flow rates. A lower bound on the proportion of rejected cases is estimated at 33 percent in this study.

Considerable research remains to be done in the area of freeway weaving sections design and analysis. Four recommendations follow.

- Fundamental work on vehicle dynamics in freeway weaving sections is needed. The procedures described in this paper are primarily empirical (data based) and do not capture the essence of vehicle interaction and its impact on average weaving and nonweaving speeds. Microscopic simulation modeling, using INTRAS *(12)* or a similar package is recommended as a cost-effective tool for conducting such analyses.
- A persistent problem throughout this study was the inadequate sample size of new (post-1970) field data. The reliability of empirical procedures can be greatly enhanced with additional data points for both calibration and validation purposes.
- Although the weaving procedure proposed in this paper has yielded superior results compared with the other three procedures, it is recommended that all four procedures be tested to solve the same problem. The final design decision must still rest with the engineer, who may select the procedure yielding the most conservative design, average out all the results, and so forth. The interactive, microcomputer program developed in this study greatly simplifies this task *(14)*.
- There is great need to tie in the safety characteristics of weaving sections (i.e., accident frequencies, type, location, and so forth) to the design and operations analysis procedures. This may result in defining lower bounds on section length and number of lanes based on accident experience.

ACKNOWLEDGMENTS

This study was sponsored by a grant from the National Highway Institute. The authors wish to thank Guido Radelat, contract manager, for his guidance and assistance throughout the study. The conclusions and recommendations presented here reflect solely the opinions of the authors and not necessarily the opinions of the National Highway Institute or the Federal Highway Administration.

REFERENCES

1. *Highway Capacity Manual*. U.S. Government Printing Office, Washington, D.C., 1950, pp. 105–116.
2. O. Norman. Operation of Weaving Areas. *HRB Bulletin 167*. HRB,

National Research Council, Washington, D.C., 1957, pp. 38–41.
3. *HRB Special Report 87: Highway Capacity Manual.* HRB, National Research Council, Washington, D.C., 1965, pp. 160–186.
4. *Weaving Area Operations Study.* NCHRP Project 3-15, Final Report. Department of Transportation Planning and Engineering, Polytechnic Institute of Brooklyn, Brooklyn, New York, 1971, unpublished.
5. *Transportation Research Circular 212: Interim Materials on Highway Capacity.* TRB, National Research Council, Washington, D.C., Jan. 1980, pp. 189–208.
6. J. E. Leisch. A New Technique for Design and Analysis of Weaving Sections on Freeways. *ITE Journal,* Vol. 49, No. 3, March 1979.
7. J. E. Leisch and J. P. Leisch. *Procedure for Analysis and Design of Weaving Sections.* Report FHWA-RD-82/54, FHWA, U.S. Department of Transportation, 1982.
8. W. Reilly, H. Kell, and P. Johnson. *Weaving Analysis Procedures for the New Highway Capacity Manual.* JHK and Associates, Aug. 1984.
9. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985, Chapter 4, pp. 4-1 to 4-19.
10. J. Fazio. *Development and Testing of A Weaving Operational Analysis and Design Procedure.* M.S. thesis. University of Illinois at Chicago, Chicago, 1985.
11. J. Fazio et al. *Users Guide to the Microcomputer Program Version of Four Weaving Operational Analysis and Design Procedures.* 2nd ed. University of Illinois at Chicago, Chicago, Aug. 1985.
12. A. G. Bullen and P. Athol. *Development and Testing of INTRAS, A Microscopic Freeway Simulation Model.* University of Pittsburgh, Pittsburgh, Feb. 1976, Vol. 2.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

A Comparison of the 1985 *Highway Capacity Manual* and the Signal Operations Analysis Package 84

DANE ISMART

The primary objective of this paper is to determine if the signalized intersection procedure as described in Chapter 9 of the 1985 *Highway Capacity Manual* (HCM) will give results consistent with the microcomputer version of the Signalized Operations Analysis Package 84 (SOAP 84). Each procedure was used to analyze the intersection in Chapter 9, Calculation 3, of the 1985 HCM. Average stopped delay was calculated for the intersection by each method and was used as the basis for comparing the 1985 HCM and SOAP 84. For through movements and protected–restricted left turns, the two procedures produced similar results for calculating stop delay, X ratios, and effective green ratios. However, for the results to be consistent, the saturation flow as calculated by the HCM method must be used in SOAP 84 as the capacity (saturation flow) for through movements and the protected–restricted left turns. For protected–permissive and unprotected left turns, the two methods produce significantly different results.

Federal Highway Administration, U.S. Department of Transportation, HHP-21, 400 Seventh Street, S.W., Washington, D.C. 20590.

Described is an effort to compare the microcomputer version of SOAP 84 with the methodology in Chapter 9, Signalized Intersections, of the 1985 *Highway Capacity Manual* (HCM).

The Signal Operations Analysis Package (SOAP 84) is a computerized method for developing control plans and evaluating the operations of individual signalized intersections. As the basis for the comparison between SOAP 84 and the 1985 HCM, delay will be calculated by each method. SOAP 84 determines average delay, which includes delay incurred during deceleration and acceleration as well as stop delay. The 1985 HCM calculates average stop delay as the basis for determining level of service. To make a comparison between the two methods, average delay will be converted to average stop delay by using the following formula *(1)*:

$$\text{Average delay}/1.3 = \text{average stop delay} \quad (1)$$

As the first step in the analysis, Calculation 3: Operational Analysis of a Multiphase Actuated Signal from Chapter 9 of

the 1985 HCM was chosen as the intersection problem to be used for the comparison. Calculation 3 was selected because it has protected permissive left turns for the north–south approaches and unprotected left turns for east–west approaches. The algorithms used by the HCM and SOAP 84 for Calculation 3 will be evaluated and compared. Because the algorithms used by the HCM and SOAP 84 will not vary for other intersections, the conclusions drawn will be valid whether 1 example or 10 are used. The worksheet for Calculation 3 is shown in Figure 1.

The second step was to develop the saturation flows and adjust volumes so that they are consistent between the two methodologies. SOAP 84 does not include saturation flow adjustment factors (lane width, grade, parking, bus blockage, area type, and right turns), which are incorporated in the 1985 HCM. Therefore, to maintain consistency the north, south, east, and west through-movement adjusted saturation flows calculated from the HCM worksheet (Figure 2) were used as the capacity input (saturation flow) for SOAP 84.

The HCM saturation flow for left turns includes a left-turn factor to account for these movements' not being able to be made at the same saturation flow rates as through movements. In the SOAP 84 program, unprotected left-turn saturation flow will be calculated based on the following equations *(2)*.

Single lane opposing flow:

$$S_L = 1{,}404 - 1.632\ V_o + .0008347\ V_o^2 - .0000002138\ V_o^3 \quad (2)$$

Multiple lane opposing flow:

$$S_L = 1{,}393 - 1.734\ V_o + .0009173\ V_o^2 - .0000001955\ V_o^3 \quad (3)$$

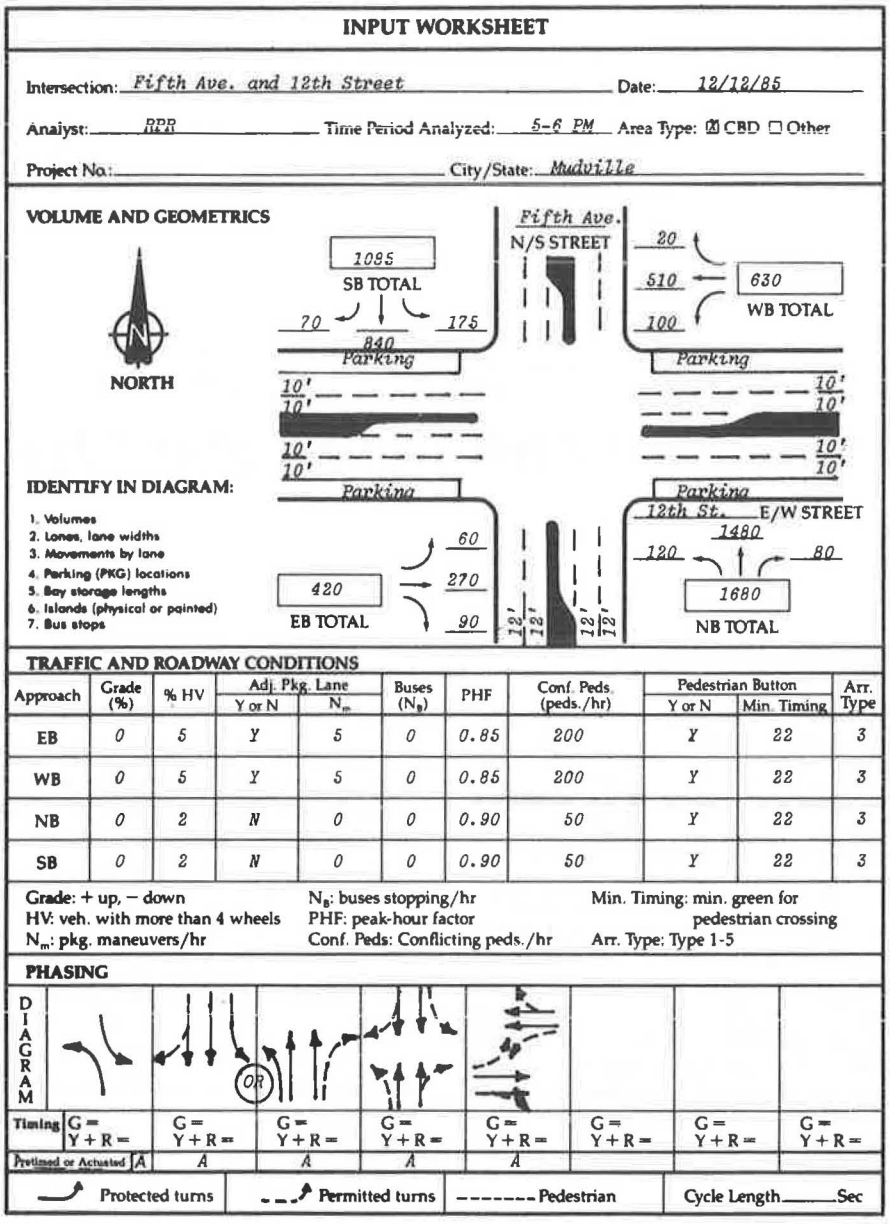


FIGURE 1 Input module worksheet for Calculation 3 in Chapter 9 of the 1985 HCM (*1*,p.9–50, Figure 9–26).

SATURATION FLOW ADJUSTMENT WORKSHEET

LANE GROUPS				ADJUSTMENT FACTORS								
① Appr.	② Lane Group Movements	③ Ideal Sat. Flow (pcphgpl)	④ No. of Lanes N	⑤ Lane Width f_w Table 9-5	⑥ Heavy Veh f_{HV} Table 9-6	⑦ Grade f_g Table 9-7	⑧ Pkg. f_p Table 9-8	⑨ Bus Blockage f_{bb} Table 9-9	⑩ Area Type f_a Table 9-10	⑪ Right Turn f_{RT} Table 9-11	⑫ Left Turn f_{LT} Table 9-12	⑬ Adj. Sat. Flow Rate s (vphg)
EB		1800	1	.93	.975	1.00	1.00	1.00	.90	1.00	.31	455
		1800	2	.93	.975	1.00	.935	1.00	.90	.94	1.00	2582
WB		1800	1	.93	.975	1.00	1.00	1.00	.90	1.00	.48	705
		1800	2	.93	.975	1.00	.935	1.00	.90	.99	1.00	2719
NB		1800	1	1.00	.99	1.00	1.00	1.00	.90	1.00	.95	1524
		1800	2	1.00	.99	1.00	1.00	1.00	.90	.99	1.00	3176
SB		1800	1	1.00	.99	1.00	1.00	1.00	.90	1.00	.95	1524
		1800	2	1.00	.99	1.00	1.00	1.00	.90	.99	1.00	3176

FIGURE 2 Saturation flow rate module worksheet for Calculation 3 in Chapter 9 of the 1985 HCM (*1*,p.9–52, Figure 9–28).

where S_L is the saturation flow for unprotected left turns [vehicles per hour (vph)], and V_o is the opposing through volume (vph).

For the protected portion of left-turn phases, SOAP 84 will use the HCM saturation flow rate without the left-turn factor (Table 1). SOAP 84 left-turn saturation flow for the unprotected portion of the turn will be based on Equation 3. Because the eastbound and westbound left turns have no protected phase, the saturation flow rates from Table 1 for eastbound left turns and westbound left turns will not be used in the calculations for total left-turn capacity of SOAP 84.

In the 1985 HCM, the volume is also adjusted based on peak-hour and lane-use factors. SOAP 84 does not make volume adjustments based on these factors. Rather, SOAP 84 relies on evaluating intersections at 15-min intervals if the user desires. With 15-min analysis periods, the peak-hour factor would be accounted for. In this example, a 1-hr time period is analyzed.

TABLE 1 SATURATION FLOW RATE WITHOUT LEFT-TURN FACTOR

Direction	Saturation Flow Rate
Eastbound left	1,469
Westbound left	1,469
Northbound left	1,603
Southbound left	1,603

Note: In SOAP 84, capacity input is specified in terms of saturation flow (vehicles per hour of green).

To remain consistent, the HCM-adjusted flow shown in Figure 3 (adjusted for peak-hour and lane-use adjustments) will be used as the SOAP volume because the SOAP analysis will be for 1 hr rather than 15-min intervals.

VOLUME ADJUSTMENT WORKSHEET

① Appr.	② Mvt.	③ Mvt. Volume (vph)	④ Peak Hour Factor PHF	⑤ Flow Rate v_p (vph) ③ ÷ ④	⑥ Lane Group	⑦ Flow rate in Lane Group v_g (vph)	⑧ Number of Lanes N	⑨ Lane Utilization Factor U Table 9-4	⑩ Adj. Flow v (vph) ⑦ × ⑨	⑪ Prop. of LT or RT P_{LT} or P_{RT}
EB	LT	60	0.85	71		71	1	1.00	71	1.0 LT
	TH	270	0.85	318		424	2	1.05	445	0.25 RT
	RT	90	0.85	106						
WB	LT	100	0.85	118		118	1	1.00	118	1.0 LT
	TH	510	0.85	600		624	2	1.05	655	0.04 RT
	RT	20	0.85	24						
NB	LT	120	0.90	133		133	1	1.00	133	1.0 LT
	TH	1460	0.90	1644		1733	2	1.05	1820	0.05 RT
	RT	80	0.90	89						
SB	LT	175	0.90	194		194	1	1.00	194	1.0 LT
	TH	840	0.90	933		1011	2	1.05	1062	0.08 RT
	RT	70	0.90	78						

FIGURE 3 Volume adjustment module worksheet for Calculation 3 in Chapter 9 of the 1985 HCM (*1*,p.9–51, Figure 9–27).

The remaining input (minimum green, headway time, phasing, and permissive left turns) needed to run SOAP 84 does not require any adjustments to remain consistent with the HCM procedure. Timing for the intersection is identified in SOAP 84 approximately as shown in Calculation 3 of the HCM. For this problem, the phasing and cycle length evaluated by SOAP 84 are given in Table 2.

TABLE 2 PHASING AND CYCLE LENGTH EVALUATED BY SOAP 84

Phase	Green Time (sec)
1	7.5
2	6.5
3	73.2
4	31.6

Note: cycle length = 118.8 sec and lost time = 1.5 sec.

By using the HCM methodology and SOAP 84, the intersection was evaluated. The results are shown in Figures 4 and 5. By using Equation 1, the delay calculated by SOAP will be converted to average stop delay. A comparison of the results given in Table 3.

The delay for through movements compares well. However, the permissive left turns are significantly different. For example, for northbound left turns, according to the HCM, delay is calculated at 71.36 sec, whereas with SOAP 84 a delay of 35.5 sec is calculated.

The first step in determining why there is such a significant difference is to look at each method's equation for determining delay. The delay equation of the HCM is as follows.

Uniform arrivals:

$$d_1 = 0.38\ C(1 - g/C)^2/[1 - (g/C)(x)]$$

LEVEL-OF-SERVICE WORKSHEET

Lane Group		First Term Delay				Second Term Delay				Total Delay & LOS		
① Appr.	② Lane Group Move-ments	③ v/c Ratio X	④ Green Ratio g/C	⑤ Cycle Length C (sec)	⑥ Delay d_1 (sec/veh)	⑦ Lane Group Capacity c (vph)	⑧ Delay d_2 (sec/veh)	⑨ Progression Factor PF Table 9-13	⑩ Lane Group Delay (sec/veh) (⑤+⑧) × ⑨	⑪ Lane Group LOS Table 9-1	⑫ Approach Delay (sec/veh)	⑬ Appr LOS Table 9-1
EB		.612	.254	118.8	29.74	116	6.27	1.00	36.01	D	28.6	D
		.678	.254	118.8	30.34	648	1.97	0.85	27.46	D		
WB		.659	.254	118.8	30.61	181	5.83	1.00	35.99	D	41.1	E
		.948	.254	118.8	33.08	682	16.33	0.85	42.00	E		
NB		.936	.653	118.8	13.94	78	57.54	1.00	71.36	F	24.5	C
		.950	.603	118.8	16.65	1915	8.14	0.85	21.07	C		
SB		.944	.690	118.8	12.17	142	42.43	1.00	54.60	E	14.7	C
		.518	.645	118.8	8.54	2049	0.19	0.85	7.42	B		

Intersection Delay 25.1 sec/veh Intersection LOS D (Table 9-1)

FIGURE 4 Level-of-service module worksheet for Calculation 3 in Chapter 9 of the 1985 HCM (*1*,p.9–56, Figure 9–31).

TABLE NO. 31
CALCULATED EFFECTIVE GREEN/CYCLE RATIO FOR EACH MOVEMENT (INCLUDING LEFT TURN RELEASE ADJUSTMENT)

*LAMDA *	TIME *	1 - NBT *	2 - NBL *	3 - SBT *	4 - SBL *	5 - EBT *	6 - EBL *	7 - WBT *	8 - WBL *
* 1 *	1500 *	.604 *	.182 *	.658 *	.152 *	.253 *	.047 *	.253 *	.085 *

TABLE NO. 33
CALCULATED DEGREE OF SATURATION (VOLUME/CAPACITY) (IF X = 999.999, NO GREEN TIME)

* X *	TIME *	1 - NBT *	2 - NBL *	3 - SBT *	4 - SBL *	5 - EBT *	6 - EBL *	7 - WBT *	8 - WBL *
* 1 *	1500 *	.949 *	.466 *	.508 *	.820 *	.680 *	1.037 *	.951 *	.955 *

TABLE NO. 38
AVERAGE UNIT DELAY (SECONDS/VEHICLE)

*AVDEL *	TIME *	1 - NBT *	2 - NBL *	3 - SBT *	4 - SBL *	5 - EBT *	6 - EBL *	7 - WBT *	8 - WBL *
* 1 *	1500 *	29.935 *	46.152 *	10.867 *	65.156 *	42.916 *	186.784 *	63.887 *	125.898 *

*** NOTE ... CYCLE LENGTH COMPUTED FROM TIMING CARD.

FIGURE 5 Calculation 3 SOAP 84 results.

TABLE 3 COMPARISON OF AVERAGE STOP DELAY PER VEHICLE CALCULATED BY USING HCM METHODOLOGY AND SOAP 84

Direction	HCM[a]	SOAP 84
Northbound through	24.79	23.02
Northbound left	71.36	35.50
Southbound through	8.73	8.36
Southbound left	54.60	50.12
Eastbound through	32.31	33.01
Eastbound left	36.01	143.68
Westbound through	49.41	49.14
Westbound left	35.99	96.84

[a]It should be noted that the HCM through movement does not include a .85 progression factor because in SOAP 84 a progression factor of 1.0 is assumed.

Random arrivals

$$d_2 = 173x^2 \{(x - 1) + [(x - 1)^2 + (16x/c)]^{1/2}\}$$

$$d = d_1 + d_2$$

where

- d = average stop delay per vehicle for the subject lane group (sec/veh)
- C = cycle length (sec),
- g/C = green ratio for the subject lane group—the ratio of effective green time to cycle length,
- x = volume-to-capacity (v/c) ratio for the subject lane group, and
- c = capacity of the subject lane group.

The delay equation of SOAP 84 is $D = D_1 + D_2 + D_3$

where

- D = average delay per vehicle (sec/veh),
- D_1 = delay per vehicle for uniform vehicle arrivals, and
- $D_2 + D_3$ = delay per vehicle for random vehicle arrivals.

Uniform arrivals:

$$D_1 = C(1 - \lambda)^2/2(1 - \lambda x)$$

where

- D_1 = delay per vehicle (sec),
- C = cycle length (sec),
- λ = proportion of green time given to the movement (effective green time/cycle length), and
- x = v/c ratio.

Random arrivals:

$$D_2 + D_3 = [(B_n^2/B_d) + (x^2/B_d)]\ 1/2 - (B_n/B_d)$$

where

- $B_n = 2(1 - x) + xz$;
- $z = (2x/v) \times (60/T) = (2/c) \times (60/T)$;
- v = approach volume (vph);
- T = period length (min), usually 60 min;
- c = capacity; and
- $B_d = 4z - z^2$.

The two factors that both equations use are the degree of saturation (x) and the effective green/cycle length. By running both delay equations with the same degrees of saturation and effective green ratios, a comparison was made to determine if the delay equations will produce different results. Figure 6 shows the delay estimates from the models.

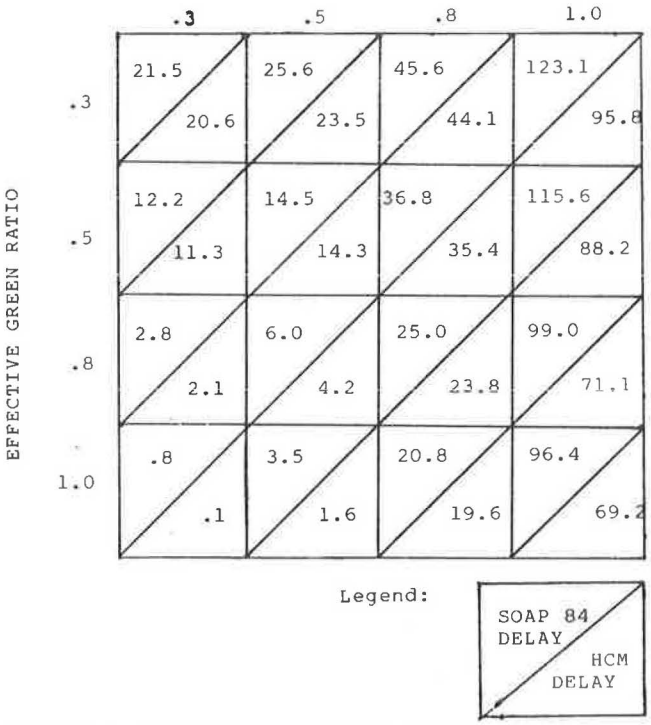


FIGURE 6 SOAP 84 versus HCM delay (x = v/c, for c = 100 sec and v = 100 vehicles).

As can be observed from the data in Figure 6, when the effective green and the v/c ratio are the same in the HCM and SOAP 84 models, the resultant delays are similar except when the v/c ratio approaches 1.0. Because the delay equations give similar results, SOAP 84 and HCM must compute the v/c and effective green ratios differently. Only by having different inputs would the two methods produce different delay estimates for the same problem, as indicated by the data in Table 3.

The next step is to look at how each method calculates the v/c and effective green ratio for left turns in Calculation 3 in Chapter 9 of the HCM. Addressed first are v/c ratios for protected–permissive left turns in the north and south directions. In the HCM methodology, the v/c ratios for the protected portion of a protected–permissive left-turn phase are based on

an arbitrary split of demand between the protected and permissive portion of the turn phase.

In Calculation 3, the HCM methodology does not assign any vehicles turning on the permissive portion of the left-turn phase. Only two vehicles per cycle are assumed to turn on the change interval (yellow) of the phase. As indicated in the HCM, a minimum of two vehicles per cycle would be turning, probably as sneakers, during the yellow phase. This assumption of the HCM is conservative because there is excess left-turn capacity in Calculation 3 for the permissive portion of the northbound left-turn and southbound left-turn phases. Because only a minimum amount of demand is assigned to the permissive portion of the left-turn phase, a high amount of left-turn volume remains on the protected left-turn portion of the phase. Thus, the v/c ratio for the protected left turn remains high and the HCM methodology uses the v/c ratio for the protected portion of a protected–permissive left turn in the delay equations.

The approximation of using the v/c ratio computed for the protected portion of the phase to represent the northbound left turns results in an excessive delay computation. The capacity for northbound left turns using the HCM equations can be approximated as follows (HCM Equation 9-22):

$$CLT = (1{,}400 - Vo)\,(g/C)\,PTL$$

where

CLT = capacity of the left-turn permissive phase, (vph),
Vo = opposing through plus right-turn movement (vph), and
(g/C) PTL = effective unsaturated green ratio for the permissive left-turn phase (sec/sec).

For NBL turns,

$$Vo = 1{,}011$$
$$(g/C)\,PTL = .45$$
$$CLT = (1{,}400 - 1{,}011) \times (.45) = 175$$

	Capacity (vph)
Protected phase	76
Permissive phase	175
Total	251

With a capacity of more than 250 vph and an adjusted demand of 133, the northbound left-turn phase will not be near saturation, as assumed in Calculation 3 in Chapter 9 of the HCM.

For the southbound direction, the HCM assumption that the protected and permissive portion of the left turns are operating at a high v/c ratio actually reflects the existing operating conditions. The permissive portion of the southbound left turns has a small amount of capacity and the delay estimates for the southbound left turns from the two methodologies are much closer (Table 3).

In the east–west direction, there is no protected left-turn phase. Left turns are made in gaps of the opposing through movements. In the HCM and SOAP 84 methodologies, the volume demand for left turns is the same for permissive-only left turns. However, the two methods do vary in their calculation of capacity for permissive left turns. The HCM methodology for estimating capacity will produce a higher value than will SOAP 84. Although the methodologies produce capacity estimates that may differ by only 40 or 50 vph for the permissive turns, they produce capacity estimates for the permissive turns with significant differences in the v/c ratios. The HCM technique with a high capacity estimates a lower v/c ratio than does SOAP 84. In Calculation 3 of the HCM, SOAP 84 estimates the v/c ratio of eastbound left turns as greater than 1.0. As a result, delay for eastbound left turns is extremely high, as estimated by SOAP 84.

Applying the HCM and SOAP 84 methodologies to Calculation 3 produces the v/c ratios given in Table 4.

As can be observed from the data in Table 4, the SOAP 84 lower estimate of capacity for unprotected left turns produces significantly different v/c ratios for eastbound and westbound left turns. For the northbound and southbound left turns, the HCM procedure of assigning a minimum of two sneakers to the permissive portion of a phase results in a higher v/c ratio than estimated by SOAP 84. Through-movement v/c ratios are almost exactly the same.

TABLE 4 V/C RATIOS ACHIEVED BY APPLYING THE HCM AND SOAP 84 METHODOLOGIES TO CALCULATION 3 IN CHAPTER 9 IN THE HCM

Direction	HCM	SOAP 84
Northbound through	0.950	0.949
Northbound left	0.936	0.466
Southbound through	0.518	0.508
Southbound left	0.944	0.820
Eastbound through	0.678	0.680
Eastbound left	0.612	1.037
Westbound through	0.948	0.951
Westbound left	0.659	0.955

The next step in the analysis is to look at effective green ratios. For left turns with permissive movements, HCM determines the effective green ratio by adding the protected (if any) and permissive phases, subtracting lost time, and dividing the result by the cycle length. SOAP 84 differs significantly in its calculation of effective green ratios for protected–permissive (northbound left and southbound left) and unprotected left turns (eastbound left and southbound left). The SOAP 84 equation is as follows:

$$\text{Left-turn effective green ratio} = [(G_p + 2.5 \times S_n) + (G_u \times C_t)/S_f]/\text{cycle length} \quad (4)$$

where

G_p = effective green time for protected portion of left turn,

S_n = number of sneakers per cycle,
G_u = unsaturated green time for permissive portion of left turn,
C_t = Tanner's capacity for unprotected left turn, and
S_f = adjusted saturation flow for protected left turn.

When the left-turn headway time is equal to 2.5, SOAP Equation 4 for left-turn effective green ratios becomes equivalent to the left-turn capacity (protected, permissive, and sneakers) divided by the saturation flow [3,600/left-turn headway (default 2.5 sec)]. The HCM includes the entire permissive green time in its estimate of effective green time for left turns. From Equation 4 only a portion of the permissive green is included in the SOAP calculation of effective green ratios. As a result, the effective green ratio for unprotected left turns will vary significantly between the two procedures.

For Calculation 3 in Chapter 9 of the HCM, the effective green ratios are given in Table 5. For through movements and protected left turns without permissive left turns, the HCM and SOAP 84 calculate effective green ratios in a similar manner:

[(Green time and clearance time) – lost time]/cycle length

Therefore, as can be observed from the data given in Table 5, all effective green ratios for through movements are the same for both methods.

TABLE 5 EFFECTIVE GREEN RATIOS

Direction	HCM	SOAP 84
Northbound through	.603	.602
Northbound left	.653	.182
Southbound through	.645	.658
Southbound left	.690	.152
Eastbound through	.254	.253
Eastbound left	.254	.047
Westbound through	.254	.253
Westbound left	.254	.083

CONCLUSIONS

Four major conclusions can be made in a comparison of the 1985 HCM and SOAP 84.

1. SOAP 84 does not make any adjustments to the capacity (saturation flow) for factors included in the HCM. Under the current version of SOAP 84, the user must estimate externally from the program the capacity (saturation flow) for each movement. Requiring the user to make an estimation of saturation flow is one of the major weaknesses in SOAP 84. This problem could be avoided if the saturation flow adjustment factors were incorporated in SOAP 84.
2. SOAP 84 and the 1985 HCM can produce similar results when estimating delay, v/c ratios, and effective green ratios for through movements and protected–restricted left turns. If 1985 HCM saturation flow adjustments are used as input, SOAP 84 could be used as a surrogate for the signalized intersection chapter of the 1985 HCM when evaluating through movements and protected–restricted left turns. However, for unprotected left turns and protected–permissive left-turn phasing, the approach taken by the two procedures differs significantly and would not give comparable results unless the procedure's algorithms are modified.
3. The 1985 HCM underassigns the number of vehicles that use the permissive phase of a protected–permissive left turn. In Calculation 3 in Chapter 9 of the HCM, no left-turning vehicles are assigned to the permissive portion of the left-turn phase. Consequently, the protected portion of the left turn is overassigned. This condition creates unrealistically high v/c ratios and delay computations.

A more realistic estimate of delay for protected–permissive left turns would result if the combined v/c ratio for the protected and permissive portion of the left turn were used in the delay computations rather than only the protected portion of the phase.

4. For unprotected left turns, the HCM procedure calculates capacity at a significantly higher value than does SOAP 84.

REFERENCES

1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
2. *SOAP 84 User's Manual, Implementation Package.* FHWA-1p-85-7. FHWA, U.S. Department of Transportation, Jan. 1985.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Entering Headway at Signalized Intersections in a Small Metropolitan Area

J. LEE AND R. L. CHEN

The entering headway is a parameter of fundamental importance to traffic engineers. It has major applications in intersection capacity and signal timing. However, the attention given to this matter appears to be inadequate. It was indicated by a literature review that past efforts tended to be infrequent, fragmented, and limited in scope. No studies were found using data from small cities or investigating factors that affect entering headways. This study, aimed at measuring entering headways in a small city and examining six factors, was conducted on sites in Lawrence, Kansas. Entering headway values from a total of 1,899 traffic queues were recorded by using video camera equipment. From the data, mean entry headways of vehicles 1 through 12 were found to be 3.80, 2.56, 2.35, 2.22, 2.16, 2.03, 1.97, 1.94, 1.94, 1.78, 1.64, and 1.76. Of the six factors studied, the following were also found: the signal type and the time of day have little influence on vehicular entering headways; vehicles in the inside lane of an intersection approach have lower entering headways than vehicles in the outside lane; vehicles in an intersection approach with lower speed limits have higher entering headways; vehicles in intersection approaches of streets with lower functional classifications have higher entering headways; and longer queue lengths appear to produce shorter entering headways for vehicles. However, because of data limitations, findings on the factors studied shall be viewed as only preliminary.

The entering headway, that is, the time between successive stopped vehicles entering a signalized intersection after the signal turns green, is of fundamental importance to traffic engineers. The values of several important parameters regarding signalized intersection operation—such as saturation flow rate, starting delay, and lost time—are often derivatives of measurements of the entering headway.

Greenshields et al. used a 16-mm camera to study traffic flow behavior at intersections in New York City and New Haven, Connecticut, in 1947 *(1)*. It was one of the earliest efforts in the United States to quantify vehicle flow characteristics on approaches to intersections. Entering headway was one of the major items investigated. Their data indicated that an average of 3.8 sec was necessary for the first stopped vehicle to enter the intersection after the traffic signal turned green. The successive mean headways for the following vehicles entering an intersection from a stopped queue were found to be 3.1, 2.7, 2.4, 2.2, and 2.1 sec.

Bartle et al. investigated starting delays and average time spacings of approaching vehicles at signalized Los Angeles area intersections in 1956 *(2)*. The mean releasing headway for the first vehicle studied ranged from 2.91 to 4.40 sec, and the average releasing headway of the remaining vehicles studied ranged from 0.95 to 1.63 sec at 13 intersection approaches. Furthermore, a significant difference in starting delay values was noted among approaches of different intersections and among different approaches of the same intersection. [Starting delay has been defined by Greenshields et al. as the additional delay caused by stop-and-start of vehicles due to traffic signals *(1)*].

Gerlough and Wanger completed a study on queue discharging behavior in 1967 utilizing field data collected from the Los Angeles metropolitan area *(3)*. They found that entering headways collected from the field were 3.85, 2.81, 2.51, 2.47, 2.37, 2.36, 2.40, 2.31, 2.24, 2.34, 2.29, 2.26, 2.19, 2.34, 2.38, 2.22, 2.26, 2.32, 2.31, and 2.28 sec from vehicles 1 to 20.

Carstens studied starting delays and headways with manual counts, stop watches, and time-lapse photographs in Ames, Iowa, in 1971 *(4)*. The elapsed time following the start of a green indication for various lengths of queues of stopped straight-through passenger cars to cross the stop line was: first car, 2.64 sec; second car, 5.13 sec; third car, 7.62 sec; fourth car, 9.91 sec; after the fourth car, add an additional 2.29 sec for each succeeding car.

King and Wilkinson used a manual input method to study the relative effectiveness of various signal configurations and lens sizes in dissipating queues in 1976 in Brookline, Massachusetts; San Francisco; Sacramento; and Huntington, New York *(5)*. The research indicated that except for lens size of the signal, no class of the signal configurations could be considered better than any other class in shortening the entering headway.

Lu used a time recorder and stop watches to collect left-turn discharging headways at an unprotected and a protected signalized intersection in Austin, Texas, in 1984 *(6)*. The average discharging headways for a protected left-turn maneuver were found to be 2.43, 2.62, 2.10, and 2.09 sec for the first four vehicles. This study should not be compared directly with other studies reviewed because it deals only with left-turning vehicles. However, it is interesting to note that left-turn vehicles were observed to have lower entering headway values.

An overall review of the studies suggests that past efforts were fragmented in terms of study methods, location characteristics, and technical objectives. Therefore, comparison of these studies is limited to a general observation that their results do not entirely agree. It is also noticed that by far the most comprehensive study was that reported by Greenshield et al. conducted 40 years ago; it is questionable whether the results could represent current traffic characteristics. In addition, the limited depth of the studies does not allow a clear identification of factors that can affect entering headway values. Furthermore, most of the studies were conducted in relatively large cities; whether the values can be applied to smaller cities remains a question to the engineer concerned with small-city traffic problems.

J. Lee, Transportation Center, Room 2013, Learned Hall, University of Kansas, Lawrence, Kans. 66045. R. L. Chen, Keith & Schnars, P.A., 1115 N.E. 4th Ave., Fort Lauderdale, Fla. 33304.

OBJECTIVE AND SCOPE OF THIS STUDY

To address some of the issues discussed earlier, the main objective of the study was an attempt to measure entering headways at signalized intersections in a small city. It was also intended to collect as much information as possible so that major factors affecting entering headways could be identified. However, because of limitations on data collection sites and their associated condition variables, this part of the study was limited to examining only a few selected factors, including signal types (actuated versus pretimed signals); time of day (morning peak versus afternoon peak); lane (inside lane versus outside lane); approach speed (speed limits of various ranges); types of street (major versus minor arterials); and queue lengths.

DATA COLLECTION AND REDUCTION

All data collection sites were located in Lawrence, Kansas. Lawrence, with a population of approximately 55,000, is the county seat of Douglas County and is one of the smallest Standard Metropolitan Statistical Areas (SMSA) in the United States. The current corporate limits of Lawrence encompass approximately 14,000 acres and more than 70 percent of the area has been developed. The main campus of the University of Kansas with an enrollment of 24,500 is located in Lawrence. Some 30,000 automobiles and 12,000 trucks are registered in Douglas County.

Sixteen signalized intersections along four major streets (Iowa Street, 23rd Street, 6th Street and Massachusetts Street) in Lawrence were selected for study. These intersections and their associated characteristics examined in this study are summarized in Table 1. Massachusetts Street between 6th and 11th Streets is located in the central business district of Lawrence and has angle parking on both sides. Both 6th and 23rd Streets are similar; they are two of the major east–west thoroughfares in Lawrence and have extensive commercial development on the roadsides. Both are four-lane streets with no parking and most of the intersections studied have protected left-turn lanes. Iowa Street is a semi-controlled access north–south major arterial in Lawrence with four lanes for traffic and protected left-turn lanes.

For all selected intersection approaches, vehicle movements were recorded by using a portable video camera system that has a built-in timer with 0.1-sec accuracy. All field videotaping of traffic movements was conducted from June to September 1984. The actual filming at each intersection included 2 hr each day with 1 hr each for a.m. and p.m. peak hours. Within the filming period, half an hour was spent on one approach and the other half for the opposite approach of the major streets.

In all, more than 32 hr of traffic data were recorded on videotapes. They represent records of approximately 5,000 single-lane traffic platoons entering the intersections. The tapes were first examined in the laboratory to screen out the cases that were not suitable for this study, including the following: platoons within which vehicles did not stop before entering an intersection; platoons with trucks; platoons with turning vehicles; and platoons in which the movements of cars were impeded by pedestrians, cross traffic, or turning vehicles. In other words, only platoons containing unimpeded, straight-through passenger cars stopped before entering an intersection

TABLE 1 DATA COLLECTION SITE CHARACTERISTICS

Intersection	Signal Type	Approach	Speed Limit	Lane
Harvard & Iowa	2-phase fully actuated	Northbound Southbound	40 mph 40 mph	Inside & Outside
15th & Iowa	4-phase fully actuated	Northbound Southbound	40 mph 40 mph	Inside & Outside
19th & Iowa	3-phase fully actuated	Northbound Southbound	40 mph 40 mph	Inside & Outside
23rd & Iowa	4-phase fully actuated	Northbound Southbound	40 mph 40 mph	Inside & Outside
27th & Iowa	3-phase fully actuated	Northbound Southbound	40 mph 40 mph	Inside & Outside
23rd & Louisiana	4-phase fixed-timed	Westbound Eastbound	35 mph 35 mph	Inside & Outside
23rd & Massachusetts	3-phase fixed-timed	Westbound Eastbound	35 mph 35 mph	Inside & Outside
23rd & Barker	3-phase fixed timed	Westbound Eastbound	35 mph 35 mph	Inside & Outside
23rd & Haskell	2-phase semi-actuated	Westbound Eastbound	45 mph 35 mph	Inside & Outside Inside & Outside
8th & Massachusetts	2-phase fix-timed	Northbound Southbound	20 mph 20 mph	Single Lane
9th & Massachusetts	2-phase fix-timed	Northbound Southbound	20 mph 20 mph	Single Lane
10th & Massachusetts	2-phase fix-timed	Northbound Southbound	20 mph 20 mph	Single Lane
19th & Massachusetts	3-phase fix-timed	Northbound Southbound	30 mph 30 mph	Inside & Outside
6th & Maine	2-phase fix-timed	Westbound Eastbound	35 mph 35 mph	Inside & Outside
6th & Michigan	2-phase semi-actuated	Westbound Eastbound	35 mph 35 mph	Inside & Outside
6th & Kasold	2-phase full actuated	Westbound Eastbound	40 mph 40 mph	Single Through

TABLE 2 STATISTICS OF ENTERING HEADWAY DATA COLLECTED

Veh	Valid Case	Mean	Std. Err	Std. Dev.	Median	Mode	Variance	Max.	Min.
1	1899	3.802	.019	.845	3.700	3.500	.714	7.800	1.600
2	1252	2.555	.018	.640	2.500	2.200	.410	5.500	1.200
3	822	2.352	.021	.612	2.300	2.100	.375	5.000	1.100
4	526	2.214	.026	.587	2.100	1.900	.345	4.400	0.900
5	327	2.163	.035	.629	2.100	1.800	.395	5.000	0.900
6	191	2.026	.040	.550	1.900	1.700	.302	4.500	1.000
7	127	1.972	.047	.527	1.900	1.600	.277	3.500	1.000
8	78	1.938	.054	.475	1.850	1.500	.225	3.900	1.100
9	44	1.941	.086	.573	1.850	1.500	.328	3.500	1.200
10	24	1.783	.074	.363	1.750	1.600	.132	2.400	1.100
11	13	1.638	.109	.393	1.600	1.300	.154	2.700	1.200
12	7	1.757	.092	.244	1.700	1.600	.060	2.100	1.500

were considered valid cases for the study. The valid cases, totaling close to 2,000 traffic queues, were later viewed on a television screen to extract the entering headway values.

For the first vehicle of a queue, its entering headway was taken to be the time elapsed between the start of a green indication and the time at which the car's rear bumper cleared the stop line. For the remaining cars in the queue, the entering headway values were taken to be the elapsed time, rear bumper to rear bumper, as the successive vehicles passed an intersection stop line.

DATA ANALYSIS AND MAJOR RESULTS

From the data reduction phase, a total of 1,899 single-lane vehicular platoons were found to be valid for this study, and the number of vehicles in the queues varied from 1 to 12. Queues with more than 12 vehicles were rare occurrences in Lawrence and were not included in the study. Entering headway values from these platoons were coded into a Honeywell 60/66 computer, and a Statistical Package for the Social Sciences was used for the statistical analyses needed.

The first step in the analysis was to derive the basic statistics out of all the entering headway data collected. The mean entering headways (in seconds) for vehicles 1 through 12 are as follows: 3.80, 2.56, 2.35, 2.22, 2.16, 2.03, 1.97, 1.94, 1.94, 1.78, 1.64, and 1.76. These and other statistics are summarized in Table 2. The first seven headways were also compared with values reported in previous studies. The comparison is shown in graphic form in Figure 1.

It is interesting to note that all the entering headways from various studies appear to follow a similar pattern except for the first two vehicles. It was suspected that studies with the first vehicle having a low entering headway may have used a different definition of entering headway than the others. The surprise, however, was that the entering headway measured in this study was lower than in almost all other studies. This was

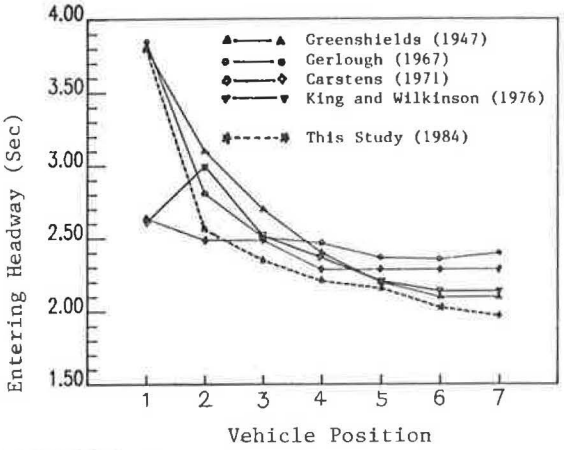


FIGURE 1 Comparison of entering headway patterns from various studies.

not expected because it was commonly assumed that small-city drivers tend to be less aggressive, which would result in higher entering headways.

A second look appeared to suggest that there are logical reasons for the observation that the entering headway measured in this study was lower than in almost all other studies. First, this study was the only one of the studies in the comparison that was conducted after the driving public felt the full impact of the 1973 energy crisis. Cars are smaller now and have better acceleration characteristics. It is also believed that in this study a large proportion of headway data may have been collected from lanes with higher approaching speeds than in the other studies. As will be revealed later, speed appears to have a significant effect on entering headways. These points not only help explain why the observed entering headways were low in this study, but also tend to suggest that entering headways can be affected by many factors, and city size may not be one of the important ones.

The second part of the analysis focused on finding out

whether several factors selected were influencing entering headways. Because the variety of conditions existing in Lawrence intersections is limited, the factors that could be examined included only signal types, time of day, lane type, approach speed, type of street, and queue length. Because most of these factors could not be expressed in continuous numerical terms, factor analysis or regression analysis could not be attempted. Therefore, a simple comparative approach was engaged.

In this approach, the factors were analyzed one at a time. For each factor analyzed, the overall headway data were grouped into categories or levels for which that particular factor is normally expressed. For example, signal types were categorized into fixed timed signals, semiactuated signals, and fully actuated signals. Thus, when analyzing the influence of signal types, the entering headways were first grouped by their associated signal types. For each group, standard statistical computations were performed to find out averages, standard deviations, and other basic statistics. After this was done, a statistical test was performed on two selected groups to see if the average entering headways of the two groups were significantly different from each other. If the difference was significant, it would be concluded that this particular factor has significant influence on entering headways.

However, the nature of the entering headway data is such that conventional statistical test methods appeared to be inappropriate. For example, assume that Signal type i has mean entering headway values for the first vehicle h_{i1}, second vehicle h_{i2}, third vehicle h_{i3}, and so on, and signal type j has mean entering headways h_{j1}, h_{j2}, h_{j3}, . . . , and so forth. Although h_{i1}, h_{i2}, h_{i3} . . . and h_{j1}, h_{j2}, h_{j3} . . . are sequential events, each mean headway within a sequence is statistically independent of other mean headways. Therefore, treating h_{i1}, h_{i2}, h_{i3} . . . as coordinates of a continuous distribution curve and comparing with a curve connecting h_{j1}, h_{j2}, h_{j3} . . . by using a χ^2-test would not be correct. On the other hand, it would also be incorrect to compare the corresponding entering headways of different groups such as $h_{i1} : h_{j1}$, $h_{i2} : h_{j2}$, $h_{i3} : h_{j3}$ . . . , and so forth and combine them by using several nonparametric pairing test techniques such as a sign test or a rank order test because one pair is actually independent of other pairs.

However, one thing that is clear is that if Group i and Group j data do not significantly differ from each other, the corresponding entering headways of different groups should not be significantly different from each other, or at least the chances of such happening are small. After consultation with mathematical statisticians at the University of Kansas, a test scheme was utilized for this study.

TABLE 3 MEAN ENTERING HEADWAY OF DIFFERENT SIGNAL TYPES AND NUMBER OF OBSERVATIONS

	TYPE OF SIGNALS								
Veh.	All fix.	All act.	2-phase fix.	3-phase fix.	4-phase fix.	2-phase full act.	2-phase semi-act.	3-phase full act.	4-phase full act.
1	3.83 (927)	3.78 (972)	3.87 (555)	3.86 (251)	3.57 (121)	3.76 (216)	3.94 (306)	3.72 (259)	3.62 (191)
2	2.61 (618)	2.50 (634)	2.66 (358)	2.52 (162)	2.57 (98)	2.47 (146)	2.63 (159)	2.56 (160)	2.36 (169)
3	2.43 (382)	2.28 (440)	2.50 (207)	2.37 (100)	2.32 (75)	2.38 (97)	2.32 (84)	2.32 (113)	2.17 (146)
4	2.29 (225)	2.16 (301)	2.37 (117)	2.14 (56)	2.26 (52)	2.07 (67)	2.09 (36)	2.26 (79)	2.16 (119)
5	2.31 (114)	2.08 (213)	2.35 (53)	2.26 (26)	2.30 (35)	2.10 (43)	2.18 (20)	2.19 (50)	2.01 (100)
6	2.10 (49)	2.00 (142)	2.20 (24)	2.09 (8)	1.97 (17)	2.01 (29)	1.93 (7)	2.09 (36)	1.96 (70)
7	2.10 (30)	1.93 (97)	2.14 (13)	2.05 (4)	2.08 (13)	1.93 (16)	1.87 (3)	2.10 (25)	1.86 (53)
8	2.13 (13)	1.90 (65)	2.30 (5)		2.03 (8)	1.78 (8)	1.50 (1)	1.97 (20)	1.90 (36)
9	2.10 (3)	1.93 (41)	2.10 (1)		2.10 (2)	2.10 (6)		1.74 (15)	2.02 (20)
10		1.78 (24)				1.68 (4)		2.03 (7)	1.68 (13)
11		1.64 (13)				1.30 (1)		1.70 (4)	1.65 (8)
12		1.76 (7)						1.90 (2)	1.70 (5)

Note: The number of observations is given in parenthesis.

TABLE 4 RESULTS OF T-TEST ON SIGNAL TYPES

Signal Type		Vehicle position in a queue											
		1	2	3	4	5	6	7	8	9	10	11	12
All fix-timed	vs. All Actuated	X	O	O	X	O	X	X	X	X	—	—	—
2-phase fix.	vs. 2-phase full act	X	O	X	O	X	X	X	X	X	—	—	—
2-phase fix.	vs. 2-phase semi-act	X	X	X	X	X	X	X	X	—	—	—	—
2-phase full act.	vs. 2-phase semi-act	X	X	X	X	X	X	X	X	—	—	—	—
3-phase fix.	vs. 3-phase full act	X	X	X	X	X	X	X	—	—	—	—	—
4-phase fix.	vs. 4-phase full act	X	X	X	X	O	X	X	X	X	—	—	—

Note: X = no significant difference; O = significant difference; and — = lack of data.

The scheme is to use a standard t-test for each corresponding entering headway pair of the groups such as h_{i1} and h_{j1}, h_{i2} and h_{j2}, h_{i3} and h_{j3} . . . , and so forth at a (10/n) percent significance interval in which n is the total number of pairs tested. The whole two groups are declared different if any pairing is shown to be statistically different. The scheme would be comparable to testing two means by using a t-test with a 10 percent level of significance.

By using the approach just described, the factors selected were analyzed. Table 3 gives a summary of entering headways grouped by signal types, and Table 4 gives the results of the t-test. Tables 5 and 6 give results of a similar analysis for a.m. and p.m. peak-hour observations. Tables 7 and 8 give the results for lane types. Tables 9 and 10 give the results for approach speeds using speed limits as representing parameters. Tables 11 and 12 give the results for different types of streets for the four streets that were involved in this study. The last analysis conducted was the effect of queue length on entering headways; the results are presented in Table 13.

Before an overall interpretation of the results was made, the information presented in Table 1 was examined. It became obvious that because of the limited variety of intersections in Lawrence, it would be difficult to draw definite conclusions about the effects of entering headways on the factors investigated in this study. For example, there are three intersections along Massachusetts Street that are different from the other study locations. They all have only one lane for approaching vehicles, have a speed limit of 20 mph, have parking on both sides, are located in a business environment, and are controlled by two-phase fixed time signals. Because of these conditions and perhaps a combination of them, the entering headways collected were significantly higher than at other sites studied. However, because a limited variety of sites existed in Lawrence, the individual impacts of the factors could not be isolated by using statistical methods. Therefore, results would be exaggerated if a particular group contained a large amount of data collected at these three intersections.

It is important to be cautious in interpreting the results of this study. With this in mind, the following summary was made.

- Signal types have no significant influence on entering headways. The apparent difference of entering headways between intersections with actuated signals and those with fixed timed signals is mostly contributed by higher headways at intersections along Massachusetts Street. Those higher headways are mostly affected by other contributing factors, notably, approach speeds.
- Time of day, signified by a.m. and p.m. peak hours, does not appear to have any influence on entering headways.
- Regarding lane type and entering headways, the inside lane of an intersection approach appeared to have slightly lower entering headways than those of the outside lane and the difference is significant. Two-lane approaches appeared to have

TABLE 5 MEAN ENTERING HEADWAY OF THE A.M. AND P.M. PERIODS AND NUMBER OF OBSERVATIONS

Veh	Time of Day	
	AM	PM
1	3.78 (950)	3.82 (949)
2	2.58 (573)	2.54 (679)
3	2.36 (344)	2.35 (478)
4	2.29 (209)	2.16 (317)
5	2.15 (122)	2.17 (205)
6	2.00 (69)	2.04 (122)
7	2.06 (45)	1.92 (82)
8	2.14 (21)	1.86 (57)
9	2.07 (9)	1.91 (35)
10	1.90 (6)	1.74 (18)
11	2.30 (2)	1.52 (11)
12	1.70 (1)	1.77 (6)

TABLE 6 RESULT OF T-TEST ON A.M. AND P.M. PEAK-HOUR PERIODS

Time of Day	Vehicle position in a queue											
	1	2	3	4	5	6	7	8	9	10	11	12
AM Peak Hour vs. PM Peak Hour	X	X	X	X	X	X	X	X	X	X	O	X

Note: X = no significant difference; O = significant difference; and – = lack of data.

TABLE 7 MEAN ENTERING HEADWAY OF DIFFERENT LANE TYPES AND NUMBER OF OBSERVATIONS

Veh	Lane Types				
	Inside Lane	Outside Lane	Single Lane	Single Through Lane	Inside & Outside Lane
1	3.71 (742)	3.76 (784)	4.10 (297)	3.95 (76)	3.74 (1526)
2	2.48 (507)	2.51 (525)	2.84 (181)	2.74 (39)	2.50 (1032)
3	2.25 (338)	2.33 (376)	2.79 (89)	2.56 (19)	2.29 (714)
4	2.09 (231)	2.23 (238)	2.76 (66)	2.12 (11)	2.16 (469)
5	2.02 (160)	2.21 (147)	3.11 (17)	2.33 (3)	2.11 (307)
6	1.98 (100)	2.04 (83)	2.57 (7)	1.70 (1)	2.01 (183)
7	2.00 (66)	1.90 (56)	2.53 (4)	1.80 (1)	1.95 (122)
8	1.84 (40)	2.02 (36)	3.40 (1)	1.40 (1)	1.93 (76)
9	1.87 (23)	2.04 (20)		1.50 (1)	1.95 (43)
10	1.68 (13)	1.95 (10)		1.40 (1)	1.80 (23)
11	1.54 (8)	1.80 (5)			1.64 (13)
12	1.70 (4)	1.83 (3)			1.76 (7)

TABLE 8 RESULTS OF T-TEST OF LANE TYPES

Lane Types	Vehicle position in a queue											
	1	2	3	4	5	6	7	8	9	10	11	12
Inside Lane vs. Outside Lane	X	X	X	O	O	X	X	X	X	X	X	X
Single Lane vs Single Through Lane	X	X	X	O	X	X	X	X	—	—	—	—
Outside Lane vs Single Through Lane	X	X	X	X	X	X	X	X	O	X	—	—
Inside & Outside Lane vs Single Lane	O	O	O	O	O	O	X	O	—	—	—	—

Note: X = no significant difference; O = significant difference; and – = lack of data.

TABLE 9 MEAN ENTERING HEADWAY OF DIFFERENT SPEED LIMITS AND NUMBER OF OBSERVATIONS

Veh.	Speed Limit of Approach				
	20 mph	30 mph	35 mph	40 mph	45 mph
1	4.10 (297)	3.89 (133)	3.76 (719)	3.70 (666)	3.77 (84)
2	2.84 (181)	2.58 (82)	2.53 (475)	2.46 (475)	2.62 (39)
3	2.79 (89)	2.46 (49)	2.30 (315)	2.28 (356)	2.29 (13)
4	2.76 (46)	2.32 (25)	2.14 (185)	2.17 (265)	1.88 (5)
5	3.11 (17)	2.31 (12)	2.14 (102)	2.07 (193)	2.67 (3)
6	2.57 (7)	1.73 (3)	2.04 (45)	2.00 (135)	1.50 (1)
7	2.53 (4)	1.90 (1)	2.03 (28)	1.93 (94)	
8	3.40 (1)		1.98 (13)	1.91 (64)	
9			2.10 (3)	1.93 (41)	
10				1.78 (24)	
11				1.64 (13)	
12				1.76 (7)	

TABLE 10 RESULTS OF T-TEST ON SPEED LIMITS

Speed Limit	Vehicle position in a queue											
	1	2	3	4	5	6	7	8	9	10	11	12
20 mph vs. 30 mph	X	O	O	O	O	X	X	—	—	—	—	—
20 mph vs. 35 mph	O	O	O	O	O	O	X	O	—	—	—	—
20 mph vs. 40 mph	O	O	O	O	O	X	X	O	—	—	—	—
20 mph vs. 45 mph	O	X	X	O	X	X	—	—	—	—	—	—
30 mph vs. 35 mph	X	X	X	X	X	X	X	—	—	—	—	—
30 mph vs. 40 mph	X	X	X	X	X	X	X	—	—	—	—	—
30 mph vs. 45 mph	X	X	X	X	X	X	—	—	—	—	—	—
35 mph vs. 40 mph	X	X	X	X	X	X	X	X	X	—	—	—
35 mph vs. 45 mph	X	X	X	X	X	X	—	—	—	—	—	—
40 mph vs. 45 mph	X	X	X	X	X	X	—	—	—	—	—	—

Note: X = no significant difference; O = significant difference; and — = lack of data.

TABLE 11 AVERAGE HEADWAY OF DIFFERENT TYPES OF STREETS AND NUMBER OF OBSERVATIONS

Veh.	Streets			
	Iowa	23rd	Mass	6th
1	3.67 (590)	3.71 (522)	4.03 (430)	3.88 (357)
2	2.43 (436)	2.53 (332)	2.76 (263)	2.59 (221)
3	2.26 (337)	2.34 (206)	2.67 (138)	2.28 (141)
4	2.17 (254)	2.17 (127)	2.60 (71)	2.06 (74)
5	2.07 (190)	2.21 (74)	2.78 (29)	2.07 (34)
6	2.01 (134)	2.01 (30)	2.32 (10)	2.04 (17)
7	1.93 (93)	2.09 (18)	2.40 (5)	1.89 (11)
8	1.91 (63)	2.00 (9)	3.40 (1)	1.84 (5)
9	1.94 (40)	2.10 (3)		1.50 (1)
10	1.80 (23)			1.40 (1)
11	1.64 (13)			
12	1.76 (7)			

significantly lower entering headway values than did approaches with only one lane. However, the data collected in this study may be biased and this statement should be regarded as preliminary.

- When approach speeds, represented by speed limits, were considered, lower speed limits in general produced higher entering headways than did higher speed limits. The difference appeared to be most significant when the lower speed limit is approximately 20 mph and the higher speed limits are above 30 mph. However, when the compared approach speed limits are all above 30 mph, the influence of speed on entering headways decreases.
- The type-of-street factor appears to be a combination of other factors studied. Massachusetts Street—because of its low speed limits, single-lane approach configurations, and roadside frictions—produced significantly higher entering headway values when compared with the other streets.
- When queue length is considered, decreasing entering

TABLE 12 RESULTS OF T-TEST ON TYPES OF STREETS

Streets Compared	Vehicle position in queue											
	1	2	3	4	5	6	7	8	9	10	11	12
Iowa St. vs. 23rd St.	X	X	X	X	X	X	X	X	X	—	—	—
Iowa St. vs. Mass. St.	O	O	O	O	O	X	X	O	—	—	—	—
Iowa St. vs. 6th St.	O	O	X	X	X	X	X	X	O	X	—	—
23rd St. vs. Mass. St.	O	O	O	O	O	X	X	O	O	—	—	—
23rd St. vs. 6th St.	O	X	X	X	X	X	X	X	X	—	—	—
Mass. St. vs. 6th St.	X	O	O	O	O	X	X	O	—	—	—	—

Note: X = no significant difference; O = significant difference; and — = lack of data.

TABLE 13 MEAN ENTERING HEADWAY OF DIFFERENT QUEUE LENGTHS AND NUMBER OF OBSERVATIONS

	Queue Length (number of vehicles)											
Veh	1	2	3	4	5	6	7	8	9	10	11	12
1	3.86 (647)	3.89 (430)	3.75 (296)	3.78 (199)	3.66 (136)	3.67 (64)	3.75 (49)	3.68 (34)	3.43 (20)	3.28 (11)	3.50 (6)	3.40 (7)
2		2.65 (430)	2.59 (296)	2.49 (199)	2.48 (136)	2.40 (64)	2.45 (49)	2.30 (34)	2.48 (20)	2.26 (11)	2.32 (6):	2.80 (7)
3			2.38 (269)	2.44 (199)	2.34 (136)	2.27 (64)	2.25 (49)	2.17 (34)	2.07 (20)	2.25 (11)	2.38 (6)	2.29 (7)
4				2.25 (199)	2.21 (136)	2.22 (64)	2.22 (49)	2.10 (34)	2.29 (20)	2.08 (11)	1.87 (6)	2.03 (7)
5					2.24 (136)	2.16 (64)	2.11 (49)	2.14 (34)	2.00 (20)	2.01 (11)	2.28 (6)	1.79 (7)
6						2.12 (64)	2.03 (49)	1.95 (34)	1.86 (20)	2.05 (11)	1.97 (6)	1.97 (7)
7							2.07 (49)	1.95 (34)	2.11 (20)	1.87 (11)	1.58 (6)	1.69 (7)
8								2.09 (34)	1.71 (20)	2.09 (11)	1.68 (6)	1.83 (7)
9									1.96 (20)	2.06 (11)	1.63 (6)	1.99 (7)
10										1.79 (11)	2.03 (6)	1.56 (7)
11											1.52 (6)	1.74 (7)
12												1.76 (7)

headway with increasing queue length appeared to be the general trend. This might have been expected because longer queues are generally associated with heavier volumes, which in turn might be associated with higher types of streets, higher speed limits, more phases in a signal setting, and a greater likelihood for drivers to be in a hurry.

CONCLUSIONS AND RECOMMENDATION

In responding to the objectives of the study, there are two major findings resulting from the study. The first finding is that from 1,899 traffic queues observed in Lawrence, Kansas, the average headways in seconds for vehicles 1 through 12 entering a signalized intersection after the light turned green were as follows: 3.80, 2.56, 2.35, 2.22, 2.16, 2.03, 1.97, 1.94, 1.94, 1.78, 1.64, and 1.76.

The second finding is that out of the six factors examined, it can be stated with some confidence that

- Signal types have little influence on entering headways at signalized intersections.
- Time of day (a.m. or p.m. traffic) has little influence on entering headways.
- The inside lane of an approach has slightly lower entering headways than does the outside lane.
- The entering headways at approaches with speed limits of 20 mph are significantly higher than those at approaches with higher speed limits ($\geq$ 30 mph). For approaches with speed limits higher than 30 mph, the influence of speed limits on the entering headway is not noticeable.
- In general, streets that have higher speed limits and less roadside frictions have lower entering headway values.
- When queue length increases, the general observation is that the entering headway values decrease.

It is to be noted that because the study locations are limited in their variety of conditions, the results on which these statements are based may be biased. Therefore, the second-part findings should be interpreted as only preliminary.

An overall review of this and previous efforts appears to indicate that although the entering headway is a basic parameter in traffic engineering, the attention received has been infrequent, fragmented, and limited in scope. All studies appear to point out that entering headways are affected by many factors. An attempt was made to examine a few of them in this study, but it appeared that the study only scratched the surface. There are numerous geometric and traffic factors that should be studied. Because the Transportation Research Board's 1985 *Highway Capacity Manual* adopts the saturation flow concept for signalized intersection capacities, a comprehensive headway study encompassing all factors appears to be urgently needed *(7)*. The ever-changing vehicle characteristics and fleet mixes on streets also indicate that the dynamics of the entering

headway require more frequent attention than has been given to this matter.

From the discussion just presented it is recommended that a systematic framework of research needs on entering headways be drawn that specifies frequency, location, and factors to be studied in a coordinated and comprehensive manner. This framework would then be made known to universities and other research institutions so that research efforts to advance the state of the art on entering headways could be guided in an organized and useful manner.

REFERENCES

1. B. D. Greenshields, D. Schapiro, and E. L. Ericksen. *Traffic Performance at Urban Street Intersections*. Eno Foundation for Highway Traffic Control, 1947.
2. R. M. Bartle, V. Skoro, and D. L. Gerlough. Starting Delay and Time Spacing of Vehicles Entering Signalized Intersection. *Bulletin 112*, HRB, National Research Council, Washington, D.C., 1956, pp. 33–41.
3. D. L. Gerlough and F. A. Wanger. *NCHRP Report 32: Improved Criteria for Traffic Signals at Individual Intersections*. HRB, National Research Council, Washington, D.C., 1967, 34 pp.
4. R. L. Carstens. Some Traffic Parameters at Signalized Intersections. *Traffic Engineering*, Vol. 41, No. 11, Aug. 1971.
5. G. F. King and M. Wilkinson. Relationship of Signal Design to Discharge Headway, Approach Capacity, and Delay. In *Transportation Research Record 615*, TRB, National Research Council, Washington, D.C., 1976, pp. 37–44.
6. Y.-J. Lu. A Study of Left-Turn Maneuver Time for Signalized Intersections. *ITE Journal*, Vol. 54, No. 10, Oct. 1984.
7. *Special Report 209: Highway Capacity Manual*. TRB, National Research Council, Washington, D.C., 1985.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Traffic Operation on Busy Two-Lane Rural Roads in The Netherlands

HEIN BOTMA

In the framework of updating standards for describing the quality of traffic flow on two-lane rural roads, research into the behavior of the traffic flow on relatively high-volume roads was carried out. Presented in this paper are findings about the relation between the volume and traffic composition as explanatory factors for speeds, headways, and platooning. It was found that mean speed was only marginally influenced by volume and truck percentage, whereas the standard deviation of speeds decreased substantially with increasing volume. An exponential tail model for headways, large enough to be relevant for passing opportunities, was used and its parameters were successfully related to volume. This model fits reality much better than the assumption that headways have a negative exponential distribution, which leads to severe underestimation of passing opportunities. Simple models were developed that relate the proportion of vehicles following in a platoon and the maximum platoon length in 5 min to volume and truck percentage. A comparison is made with results in the proposed Chapter 8 on Two-Lane Highways of the 1985 *Highway Capacity Manual*.

Transportation Research Laboratory, Department of Civil Engineering, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands.

In The Netherlands, as in many other countries, traffic engineers have used information from the 1965 *Highway Capacity Manual* (HCM) to a considerable extent. However, it was increasingly believed that the procedures and data in the HCM were not always applicable. Some factors were clearly different in the United States and The Netherlands, for example, capabilities of automobiles and trucks. Moreover, some data in the HCM appeared to be outdated, as was probably realized first in the United States.

As part of determining and updating standards for design of all types of rural roads, the Transportation and Traffic Engineering Division of the Dutch Ministry of Transport and Public Works has given attention to two-lane rural roads for motor vehicles, that is, roads without bicycles or low-speed motorbikes (mopeds) on the carriageway. Given in this paper are some results of research into the behavior of the traffic flow at relatively high volumes. This behavior is relevant for determining the service volumes that define the levels of service.

As is generally accepted, speed is no longer the only most suitable criterion on which to base level of service because speeds do not depend on volume that much any more. Speed was replaced by a general criterion that can best be described as

freedom to maneuver, a concept that was translated into operational quantities such as passing possibilities, waiting time before a passing can be carried out, proportion of vehicles in a platoon, and platoon length. Based on partly known and partly estimated relations of several quantities with volume, preliminary service volumes were chosen that should be applicable for road sections with ideal geometrics and with 100-percent person cars.

To provide a stronger base for the procedure, answers were needed to questions about assumptions, relations, and values of parameters used. Goals of the study were formulated by a committee in which several Dutch road administrators participated; this study was carried out under contract by a Dutch consultant, Bureau Goudappel Coffeng B.V. at Deventer, and the Transportation Research Laboratory of the Delft University of Technology. This paper is restricted to a part of the research carried out by the Transportation Research Laboratory and does not address any consequences of the procedure used to determine the service volumes.

RESEARCH GOALS AND METHODS

Briefly, the goals of the research are to determine the relation between volume and vehicle composition and the following: speeds, headways, and platoons. This should be done on two-lane rural roads with ideal geometrics, that is,

- Level and straight sections with sufficient sight distance over at least 0.6 mi (1 km).
- Lane width between the lines of 10.2 to 10.7 ft (3.1 to 3.25 m) and clear shoulders; [the total pavement width is 23 to 25 ft (7 to 7.5 m); hence, no paved shoulders that make easy passing maneuvers possible are present].
- No intersections or other disturbances over at least 1.2 mi (2 km).

The roads should be busy, that is, with annual average daily traffic (AADT) greater that 10,000 vehicles, and weather conditions should be good.

Consequently, the explanatory factors are volume and traffic composition. Because there are two flows on a two-lane road, it was decided to define four factors in relation to volume and traffic composition in each direction. Only two types of vehicles were considered: person cars and trucks. Although it was recognized that there is a wide range of types of trucks that differ in size and power, because road authorities have data mostly on one general truck percentage, it was not considered relevant to differentiate further.

In U.S. literature, much attention is paid to the presence of recreational vehicles. Although there are roads in The Netherlands that have a high percentage of automobiles towing caravans in certain seasons, this problem did not have sufficient importance to be addressed in this rather general research.

The four factors defined in relation to volume and traffic composition in each direction were speeds, headways, platoons, and homogeneousness. Each of these factors will be discussed further.

Speeds

It was decided to represent speeds by their mean and standard deviation. Although it was known in advance that mean speed is not influenced much by volume, it was considered worthwhile to check this point again. The standard deviation of speeds was considered to be relevant for passing demand and as a measure of homogeneousness of the speeds. It was expected that it would decrease with volume, but the size of the decrease was unknown.

Headways

Passing opportunities on a two-lane road, given the sight distance is sufficient, are determined by the distribution of headways of the opposing flow. A well-known and much-used assumption is that headways have a negative exponential distribution (NED). It is then simple to calculate the proportion of headways greater than, for example, 20 sec, that can be considered to represent passing opportunities. It is also known that the NED assumption is valid only for low volumes; however, up to which volume is it applicable and what is the error for higher volumes? The goal of the research, therefore, was to determine if the NED assumption is a useful approximation, and, if not, to indicate its error when calculating the frequency of large headways.

Platoons

When passing demand exceeds passing opportunities, vehicles will have to be driven in platoons. Consequently, the level of platooning can be considered to represent an aspect of the quality of the traffic flow as experienced by the drivers. The problems are as follows:

- Determining when a vehicle belongs to a platoon, and
- Defining simple parameters that represent the level of platooning.

Concerning the first problem, to keep things simple, it was decided to define a platoon leader as a vehicle with a headway greater that h_o sec; hence, followers in a platoon have a headway smaller than h_o. In a preliminary analysis of the data, h_o was given the values of 4, 5, and 6 sec. It was found that results with these three values were congruent and no preference could be deduced. Nor did speed differences between a follower in a platoon and his predecessor make it possible to discriminate among the three values. Finally, a value of 5 sec was chosen, in accordance with a proposal of the Organisation for Economic Cooperation and Development *(2)* and based on a detailed research of following behavior on a two-lane road by Botma et al. *(3)*. The same value was later chosen for the proposed 1985 HCM *(4)*.

Concerning the second problem, the level of platooning—as experienced by the drivers—can be well represented by the proportion of drivers that are following in a platoon. Given the platoon definition used, this proportion of hindered drivers

equals the proportion of headways less than 5 sec. To gain insight into the size of large platoons, the maximum platoon length that passed in 5 min was also analyzed.

Homogeneousness

In most research of traffic flow on homogeneous two-lane roads, it is tacitly assumed that the properties of the flow are also homogeneous. However, when passing demand exceeds passing possibilities it can be expected that, moving in the direction of the traffic stream on a lane, the speeds will decrease and the platooning will increase.

In preliminary research, this expectation was confirmed with respect to platooning; however, with respect to speed, results were inconclusive. To obtain more knowledge of this phenomenon, it was decided to collect data simultaneously at three cross sections of homogeneous road sections at distances of 0.3 mi (0.5 km).

DATA COLLECTION

Data were collected in 1984 at five road sections throughout the country, and with AADT ranging from 10,000 to 25,500 vehicles (see Table 1). Measurements were carried out simultaneously at both lanes and three cross sections 0.3 mi (0.5 km) apart. From each passing vehicle, the moments that the axles passed a cross section were registered on magnetic tape. Using these moments, the following items per vehicle can be calculated: passing moment, headway, speed, wheelbases, and vehicle type. The vehicle type is deduced from the number of axles and the distances between them. Data were collected on two days and during three periods per day, morning rush hours (7 to 9 a.m.), nonrush hours (2 to 4 p.m.), and evening rush hours (4 to 6:30 p.m.)

TABLE 1 STUDY SITE DATA

Site No.	Site Name	AADT (no. of veh/24 hr)	Average Truck Percent	Speed Limit mph	Speed Limit km/h
1	Vinkeveen	16,400	15	50	80
2	Hellegat	13,900	14	62	100
3	Grouw	10,100	12	62	100
4	Maarn	25,500	24	50	80
5	Doenkade	17,900	13	50	80

The passing of vehicle axles is detected with fairly unobtrusive, thin rubber tubes of 0.2 in. (5 mm) diameter at a distance of 3 ft (1 m) on the road. In preliminary research, it was found that only during the first 24 hr after installation did there appear to be a small decrease in speeds. Consequently, the tubes were installed at least 1 day before the measurements were carried out, giving the regular passing driver the opportunity to become accustomed to them. Most drivers probably did not notice the tubes because all other cables and the van with equipment were perfectly hidden.

ANALYSES AND RESULTS

Time Interval

To analyze the relations described in the Section Research Goals and Methods, it was necessary to divide the measurement periods of at least 2 hr into smaller time intervals. A 5-min interval was chosen as a compromise between having a rate of flow that is fairly constant and having a sufficient number of vehicles in an interval. This interval also proved to be satisfactory in a recent Canadian research on two-lane roads *(5)*. Care was taken that no platoons were cut into pieces and distributed over two neighboring intervals by shifting the boundary between two intervals slightly.

Procedure

It was found that results of the two days were not considerably different and consequently could be combined into one data set, which meant that there were 30 data sets left, from 5 locations, 3 periods, and 2 directions. Because the 5-min interval was used, at least 48 data points were available per data set. Relations were obtained separately for each data set, and combinations were made dependent on the results. This was done by using statistical testing and deciding if significant differences were relevant from a practical engineering point of view.

Technique

Linear regression was used to investigate the relations that were assumed to be mostly linear. Nonlinear relations were hypothesized only when they could be theoretically based or when the data gave firm evidence of nonlinearities.

Mean Lane Speed

The mean lane speed was related to the volume and the truck percentage in both directions. it was found that only volume and truck percentage of the same direction had a linear influence on the mean lane speed. Relations were weak, and the explained variance varied from 4 to 55 percent. Over the lane-volume range of 300 to 1,100 vehicles per hour (vph), mean speed decreased on the average 6 mph (10 km/h). This confirms the expectations that the speed-volume curve is rather flat. The range of truck percentage was 5 to 30 percent; leading to a decrease in mean speed of 3 mph (5 km/h) on the average, which is rather small.

Mean lane speed was rather inhomogeneous; changes over 0.6 mi (1 km) were significant in two-thirds of all data sets. In most cases, the mean lane speed decreased in the direction of the stream, as was expected; however, in two cases it increased. This phenomenon is not yet understood, and it can only be concluded that road sections longer than 0.6 mi (1 km) are needed to provide additional understanding. Consequently, the absolute level of mean speed is somewhat unknown, its range being 50 to 62 mph (80 to 100 km/h).

Mean Two-Way Speed

Mean two-way speed was related to the two-way volume, the two-way truck percentage, and the imbalance between the lane volumes. This imbalance was defined as (maximum lane volume/minimum lane volume) minus 1 [see, for example, Steierwald and Schmidt *(6)*].

It was found that only volume and truck percentage had a significant and linear effect on mean speed. Relations were somewhat stronger than for mean lane speed, but still rather weak: 11 to 63 percent explained variance. Over the two-way volume range of 600 to 2,000 vph, mean speed decreased on the average 7½ mph (12 km/h); over the truck percentage range of 5 to 30 percent, an average decrease of 3 mph (5 km/h) was found.

Results of mean lane speed and mean two-way speed are fairly consistent; in both cases, volume and vehicle composition have a rather small influence.

Standard Deviation of Lane Speeds (SIGMA)

It was found that the data sets had a consistent behavior and that the inhomogeneities found did not prevent the combination of all data. The total result was (in mph):

$$\underset{\substack{\\ (3.4\%)\ \ (21\%)\ \ \ \ \ (41\%)}}{\text{SIGMA} = 8.5 - 2.3\ 10^{-3} * Q_1 - 1.2\ 10^{-3} * Q_2} \qquad R^2 = 66\%\quad N = 30 \qquad (1)$$

where Q_1 is lane volume in direction of SIGMA, and Q_2 is the opposing volume. The standard errors of the coefficients are placed underneath in parentheses.

Hence, in this case, lane volumes in both directions have a significant influence on the standard deviation. However, the influence of the opposing volume is relatively small and its standard error rather high. It may be more practicable to use a model with only Q_1 as a factor, that is,

$$\underset{\substack{\\ (3.4\%)\ \ \ (15.5\%)}}{\text{SIGMA} = 8.3 - 2.9\ \ 10^{-3} * Q_1} \qquad R^2 = 60\%\quad N = 30 \qquad (2)$$

The latter equation means that over the range of lane volumes of 300 to 1,100 vph, the standard deviation decreases from approximately 7.5 to 5 mph (12 to 8 km/h), a substantial change. Moving in the direction of the flow, SIGMA decreases on the average by 7 percent over 0.6 mi (1 km), an inhomogeneousness that cannot be overlooked and can be interpreted as more platooning.

Headways

In a preliminary analysis, it was found that the NED did underestimate the frequency of large headways. As an alternative to the NED, an exponential tail (ET) model was used, taken from the family of semi-Poisson headway models [see, for example, Branston *(7)*]. In this model, it is assumed that for headways greater than a certain value T, vehicles do not interact and the distribution is exponential. The form of the distribution for shorter headways is not relevant for the frequencies of large headways. A value of 10 sec was chosen for T to ensure that the condition of no interaction was fulfilled. The probability that a headway h is greater than a fixed value h_i is as follows:

$$P_i = Pr\ (h > h_i) = P_o * \exp[-\lambda * (h_i - T)] \qquad (3)$$

Parameter P_o is the probability that headways are greater than T seconds and parameter λ determines the steepness of the tail.

P_o and λ can be estimated for each time interval and related to the explanation factors. It was found that only the lane volume of the direction considered had a significant effect and that 11 data sets could be combined. The final models were as follows:

$$\underset{\substack{\\ (5.1\%)\ \ (1.0\%)}}{\ln (P_o) = -0.286 - 2.29\ 10^{-3} * Q_1} \qquad R^2 = 86\%\quad N = 1{,}582 \qquad (4)$$

$$\underset{\substack{\\ (11.0\%)\ \ (4.1\%)}}{\lambda = 0.0314 + 0.132\ 10^{-3} * Q_1} \qquad R^2 = 27\%\quad N = 1{,}582 \qquad (5)$$

Substituting these relations into the ET model (Equation 3) leads to

$$\ln (P_i) = 0.028 - 0.97\ 10^{-3} * Q_1 - 31.4\ 10^{-3} * h_i - 0.132\ 10^{-3} * Q_1 * h_i \qquad (6)$$

The latter formula is used to determine the proportion of headways greater than 14, 17, 21, and 23 sec for each 5-min interval and to do the same using the NED. The chosen values of the headways stem from a part of the study on passing behavior that is not described in this paper.

It can be observed from the data in Table 2 that the mean difference between calculated and measured proportion is small for the ET model, and large and always negative for the NED. The standard deviation of the differences is rather high; however, it should be realized that the number of large headways in 5 min is a highly fluctuating quantity.

It is remarkable that although the relation of λ with Q_1 (Equation 5) is rather weak, the end results are satisfying. It

TABLE 2 MEAN AND STANDARD DEVIATION OF PROPORTIONAL DIFFERENCES BETWEEN CALCULATED AND MEASURED PROPORTION OF HEADWAYS GREATER THAN 14, 17, 21, AND 23 SEC ACCORDING TO EXPONENTIAL TAIL (ET) MODEL AND NEGATIVE EXPONENTIAL DISTRIBUTED (NED) HEADWAYS

Headway (sec)	Mean Differences (%)		Standard Deviation Differences (%)	
	ET	NED	ET	NED
14	1.4	−15.9	34	31
17	0.4	−26.0	42	36
21	0.0	−39.4	51	41
23	−1.2	−39.0	53	41

should be added that the ET model should not be used for lane volumes less than 300 to 400 vph, in which case preference should be given to the NED. This point is shown in Figure 1, in which the number of headways greater than 21 sec is given as a function of the lane volume according to the ET model and the NED.

To investigate the homogeneousness of the headways, the standard deviation was analyzed. It was found that it increased moving with the flow, corresponding to more large headways and larger platoons.

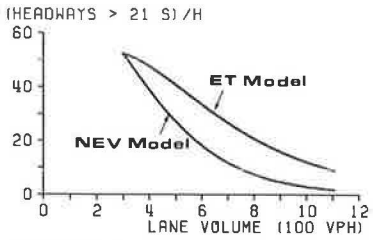


FIGURE 1 Relationship between number of headways greater than 21 sec and lane flow according to exponential tail model and negative exponential distributed headways.

Platooning

The proportion of hindered vehicles in platoons (PRHIN) and the maximum platoon length in 5 min (expressed in number of vehicles and denoted by MAXLEN) were related to the explanatory factors. It was found that the lane volume (Q_1) and the truck percentage (TP_1) of the flow considered gave the most satisfying models.

All data sets could be combined, leading to the following:

$$\underset{\substack{(1.3\%) \quad (8.5\%) \\ R^2 = 77\% \quad N = 1{,}582}}{PRHIN = 1 - \exp[-1.70\ 10^{-3} * Q_1 - 6.69\ 10^{-3} * TP_1]} \quad (7)$$

$$\underset{\substack{(2.6\%) \quad (1.9\%) \quad (25.4\%) \\ R^2 = 64\% \quad N = 1{,}582}}{MAXLEN = 2.90 * \exp[1.84\ 10^{-3} * Q_1 + 4.02\ 10^{-3} * TP_1]} \quad (8)$$

It can easily be derived that according to an NED and a platoon criterion of 5 sec, Equation 7 would have been

$$PRHIN = 1 - \exp[-5 * Q_1/3{,}600] \quad (9)$$

This can be compared with Equation 7 with TP_1 set at 0, leading to

$$PRHIN = 1 - \exp[-6.12 * Q_1/3{,}600] \quad (10)$$

Consequently, one needs to replace 5 sec with 6.12 to compensate for the facts that the NED is not valid and that the real level of platooning is higher. The inhomogeneousness of PRHIM and MAXLEN was consistent with that of the standard deviations of the speeds and headways. Moving in the direction of the flow, on the average PRHIN increased by 2.6 percent and MAXLEN by 0.7 vehicle over 0.6 mi (1 km). The extra representation of the platooning by MAXLEN does not lead to much independent information because PRHIN and MAXLEN have a correlation coefficient of 0.76, and the logarithms of both variables correlate even with 0.88.

COMPARISON WITH PROPOSED 1985 HCM

At the end of this project, the proposed Chapter 8 on Two-Lane Highways of the 1985 HCM became available *(4)*. In this proposal, two relations are given that can be compared with results of this research, that is, mean two-way speed and the percent time delay as a function of two-way volume.

In Figure 2, the relationship between average speed and volume is given according to the proposed 1985 HCM and the corresponding data found in this research, corrected to a truck percentage of zero by means of the regression equations found. It can be concluded that mean speeds on two-lane roads with ideal geometrics found in this research are 3 to 6 mph (5 to 10 km/h) higher than the average result obtained in the United States.

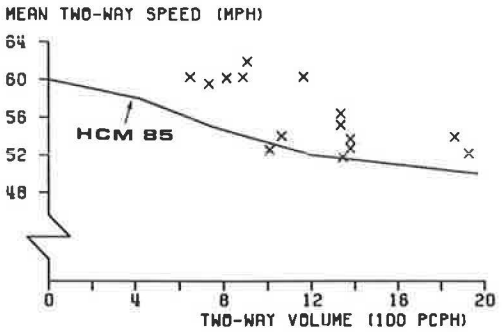


FIGURE 2 Relationship between average two-way speed and two-way volume according to the proposed 1985 HCM and the values found in this research.

The percentage time delay in the 1985 HCM is considered to be the most relevant criterion for level of service. It is operationally defined as the proportion of vehicles with headways less than 5 sec, so it is equal to the proportion of hindered vehicles (PRHIN) in this paper. Its dependence on volume is given for a two-way volume and ideal conditions, which includes a 50/50 directional split. Consequently, it can be assumed that the relation is the same for both lane flows.

Table 3 gives a comparison for volumes equal to the service volumes of the HCM. It can be concluded that the differences between both results are surprisingly small. It should be added that the model presented in this paper is not valid for volumes much greater than 1,200 vehicles per lane and hour, and that its exponential form makes it impossible to reach 100 percent hindered vehicles at capacity.

In interpreting the conclusions reached, the general conditions of this study should be kept in mind. Data were collected on two-lane rural roads with ideal geometrics and under good weather conditions; lane volumes had a range of 300 to 1,100

TABLE 3 COMPARISON OF PERCENT TIME DELAY OF PROPOSED 1985 HCM AND RESULTS OF THE MODEL IN THIS PAPER

Level of Service	Percent Time Delay (United States[a])	Volume-to-Capacity Ratio	Service Volume	Percent Time Delay (The Netherlands[b])
A	<30	0.15	420	30.0
B	<45	0.27	750	47.1
C	<60	0.43	1200	63.9
D	<75	0.64	1800	78.3
E	>75	1.00	2800	–

[a]Using the proposed 1985 *Highway Capacity Manual* (HCM).
[b]Using the results of the model in this paper.

vph and two-way volumes of 600 to 2,000 vph, and the truck percentage varied from 5 to 30 percent.

Theoretical analysis of the traffic flow process on a two-lane rural road leads to expectations that the flows in both directions are interacting and have dependent properties. However, in this rather practical research it was seldom found that opposing volume and truck percentage substantially influenced the properties of the flow considered. Part of this lack of influence may be due to dependencies between explanatory factors; lane volumes had a correlation of 0.37 with each other and lane truck percentages had a correlation of 0.36. Correlations between lane volumes and truck percentages were not greater than 0.1.

Mean lane speed decreases with lane volume and lane truck percentage; the decrease is relatively small. Mean two-way speed decreases with increasing two-way volume and two-way truck percentage. In accordance with mean lane speed, the decrease is minor. No influence of the directional split could be detected. Values found in this study are somewhat higher [3 to 6 mph (5 to 10 km/h)] than mean speeds presented in the proposed 1985 HCM, referring to both roads and traffic with practically the same ideal conditions.

Standard deviation of lane speeds decreases with lane volumes in both directions but is not dependent on truck percentage. Over a lane volume range of 300 to 1,100 vph, the standard deviation decreases from approximately 7½ to 5 mph (12 to 8 km/h).

The frequencies of large headways that are relevant for passing opportunities can be predicted by a model that needs only lane volume as an input. The model is valid for lane volumes greater than 300 to 400 vph and shows that the assumption of negative exponential distributed headways leads to a severe underestimation of the frequency of large headways.

The level of platooning can be well represented by the proportion of vehicles that have to follow in a platoon. This proportion depends on the lane volume and truck percentage. There is a good similarity between this model and the relation presented in the proposed 1985 HCM.

Traffic flow on homogeneous sections of two-lane rural roads can still be inhomogeneous, that is, properties of the flow change with the position along the road. Moving with the flow on a lane, the standard deviation of speeds decreases and that of headways increases. These two changes are consistent with larger platoons and more frequent large headways that were found. However, the phenomenon is only partly understood and requires more research over sections of roads that are substantially longer than 0.6 mi (1 km), which were used in this study.

ACKNOWLEDGMENT

The research described in this paper was sponsored by the Dutch Ministry of Transport and Public Works, more specifically, the Transportation and Traffic Engineering Division of Rijkswaterstaat. We wish to thank the sponsor for his permission to publish these results.

REFERENCES

1. *HRB Special Report 87: Highway Capacity Manual.* HRB, National Research Council, Washington, D.C., 1965, 411 pp.
2. *Speed Limits Outside Built-Up Areas.* Organisation for Economic Cooperation and Development, Paris, France, 1972.
3. H. Botma, J. H. Papendrecht, and D. Westland. *Validation of Capacity Estimators Based on the Decomposition of the Distribution of Headways.* Transportation Research Laboratory, Delft University of Technology, Delft, The Netherlands, 1980.
4. *Transportation Research Circular 284: Proposed Chapters for the 1985 Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., Oct. 1984.
5. M. van Aerde and S. Yagar. Volume Effects on Speed on Two-Lane Highways in Ontario. *Transportation Research,* Vol. 17A, No. 4, 1983, pp. 301–313.
6. G. Steierwald and G. Schmidt. Zur Problematik des Erhebungszeitraums bei Stromzählungen des individuellen Strassenverkehrs. *Strassenverkehrstechnik,* Vol. 13, No. 2, 1969, pp. 46–55.
7. D. Branston. *Model of Single Lane Traffic Flow.* Ph.D. dissertation. University College, London, England, 1974.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.